Virology 494 (2016) 198–214

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Evolution of coreceptor utilization to escape CCR5 antagonist therapy

Jie Zhang [a], Xiang Gao [a], John Martin [b], Bruce Rosa [b], Zheng Chen [c,1], Makedonka Mitreva [b], Timothy Henrich [d,2], Daniel Kuritzkes [d], Lee Ratner [a,*]

[a] Division of Molecular Oncology, Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA
[b] The McDonnelle Genome Institute, Washington University School of Medicine, St. Louis, MO, USA
[c] Department of Computer Science and Engineering, Washington University in St. Louis, St. Louis, MO, USA
[d] Division of Infectious Diseases, Brigham and Women's Hospital, Harvard Medical School, MA, USA

ARTICLE INFO

Article history:
Received 30 November 2015
Returned to author for revisions
4 April 2016
Accepted 7 April 2016
Available online 26 April 2016

Keywords:
HIV/AIDS
Tropism shift
Coreceptor inhibitor
Drug resistance

ABSTRACT

The HIV-1 envelope interacts with coreceptors CCR5 and CXCR4 in a dynamic, multi-step process, its molecular details not clearly delineated. Use of CCR5 antagonists results in tropism shift and therapeutic failure. Here we describe a novel approach using full-length patient-derived gp160 quasispecies libraries cloned into HIV-1 molecular clones, their separation based on phenotypic tropism in vitro, and deep sequencing of the resultant variants for structure-function analyses. Analysis of functionally validated envelope sequences from patients who failed CCR5 antagonist therapy revealed determinants strongly associated with coreceptor specificity, especially at the gp120-gp41 and gp41-gp41 interaction surfaces that invite future research on the roles of subunit interaction and envelope trimer stability in coreceptor usage. This study identifies important structure-function relationships in HIV-1 envelope, and demonstrates proof of concept for a new integrated analysis method that facilitates laboratory discovery of resistant mutants to aid in development of other therapeutic agents.

1. Introduction

Infection by human Immunodeficiency Virus Type 1 is initiated by viral envelope trimer binding to the host cell receptor, CD4, and subsequently one of the two coreceptors C-C Chemokine Receptor Type 5 (CCR5) or C-X-C Chemokine Receptor Type 4 (CXCR4). A series of step-wise, dynamic conformational changes occurs following the two binding events which culminates in viral and cell membrane fusion, and release of the nucleocapsid core into the host cell (Wilen et al., 2012; Berger, 1997). The ability of a virion to bind either one or both chemokine receptors for entry has significant implications in disease pathogenesis, and affects treatment options (Gorry and Ancuta, 2011). Although nearly all founder / transmitter viruses are strictly CCR5-tropic (termed R5 virus), close to 50% of HIV-1 subtype B patients will spontaneously develop CXCR4-using (X4) virus as the disease progresses, and the presence of X4 virus correlates with a worse clinical prognosis (Connor et al., 1997). This observation is thought to reflect the bottleneck nature of HIV-1 transmission, a preferential targeting of previously activated memory T lymphocytes in the early phase of infection, and a selection pressure that drives HIV-1 infection of naïve T cells late in the disease course when the majority of memory T cells are depleted (Schuitemaker et al., 2011). The process of coreceptor shift that otherwise naturally occurs with disease progression can take an expedited course upon the administration of a CCR5 antagonist antiretroviral drug (Moore and Kuritzkes, 2009). This limits the therapeutic application of CCR5 antagonists in patient management, and requires meticulous tropism determination prior to administration of the drug (Henrich and Kuritzkes, 2013; Ray and Doms, 2006).

Numerous studies examined the structural aspects of HIV-1 envelope and coreceptor interaction (Picard et al., 1997; Samson et al., 1997; Siciliano et al., 1999; Pontow and Ratner, 2001). There has been a continuing effort to determine the viral envelope protein structure in its ligand-free, mature trimer configuration, as well as in complex with neutralizing antibodies (Kwong et al., 1998; Poignard et al., 2001; Do Kwon et al., 2015). Recently, high-resolution crystal structures of CCR5 and CXCR4 in complexes with small molecular ligands, have been determined (Wu et al., 2010; Saita et al., 2006; Tan et al., 2013). These findings shed light on the molecular interaction between the envelope and coreceptor. However, because the viral entry process is highly dynamic and involves large conformational changes of a complex macromolecular conglomeration, there are still large gaps in our understanding at a molecular level. The Variable loop 3 (V3) region has been shown to play a critical role in coreceptor

* Correspondence to: Department of Medicine, Washington University School of Medicine, 660 S. Euclid, Campus Box 8069, St. Louis, MO 63110, USA.
E-mail address: lratner@dom.wustl.edu (L. Ratner).
[1] Present address: Facebook, Inc., Menlo Park, CA
[2] Present address: UCSF School of Medicine, San Francisco, CA, USA

http://dx.doi.org/10.1016/j.virol.2016.04.010

specificity, by making direct contact with the extracellular loops of CCR5 or CXCR4 (Westervelt et al., 1992). Regions outside V3 have been found to affect tropism as well, but it is less clear how they affect the entry process (Pastore et al., 2006; Cashin et al., 2014; Dobrowsky et al., 2013).

Currently, tropism determination approaches can be broadly categorized as phenotypic or genotypic approaches. Phenotypic testing, such as the Trofile assay, is primarily based on ex vivo determination of tropism in established coreceptor expressing cell lines, and has been the gold standard for tropism determination (Low et al., 2009). Recent advances in Next-Generation Sequencing (NGS) technologies have made genotypic tropism testing more feasible and potentially a cheaper, faster, and more accurate alternative to phenotypic methods (Prosperi et al., 2010; Archer et al., 2012; Sede et al., 2014; Swenson et al., 2011; Tsibris et al., 2009). The current challenge with the wide application of genotypic tropism testing lies in the ability to accurately predict tropism based on amino acid sequences. Current algorithms such as Geno2Pheno and PSSM focus exclusively on the V3 region sequence for tropism prediction, and the outcome, though accurate for CXCR4 utilization, has a high false positive rate, and limits the application of CCR5 antagonist drug to patients who could potentially benefit from the therapy (Lengauer et al., 2007; Jensen et al., 2003; Sirois et al., 2005). Therefore, it is important to better understand the molecular determinants for CXCR4 usage and the structure-function relationship of HIV-1 envelope, and use this information to enhance the predictive power of the current methods in clinical and computational studies (Cashin et al., 2014; Thielen et al., 2011). In addition, new determinants critical for CXCR4 usage can be attractive drug targets for development of future entry inhibitors.

Given the challenge that there are a limited number of phenotypically validated envelope sequences, in which most of the available sequences include only the V3 region, we devised a novel comprehensive approach that combines phenotypic tropism determination with NGS technologies, and applied it to study a panel of samples from patients who failed a CCR5 inhibitor clinical trial. By analyzing NGS data of phenotypically validated full-length HIV-1 gp160 sequences, we identified key determinants outside V3 that correlate with coreceptor use and CCR5 inhibitor resistance.

2. Materials and methods

2.1. Patient samples

HIV-1 envelope amplicons were obtained from participants in a Vicriviroc (VCV) Phase II clinical trial (AIDS Clinical Trials Group [ACTG] A5211; NCT00082498) (Gulick et al., 2007). Participant selection and sample preparation were described previously (Tsibris et al., 2009). Plasma samples were collected from study participants who experienced virologic failure and had tropism change by the phenotypic Trofile assay (Monogram Biosciences). Pelleted virion RNA was used to prepared *env* cDNA. Envelope amplicons were generated using previously reported primers (Kirchherr et al., 2007):

Env1Atopo (5′-CACCGGCTTAGGCATCTCCTATGGCAGGAAGAA-3′)
FLenv2.2 (5′-AGCTGGATCCGTCTCGAGATACTGCTCCCACCC-3′)

2.2. Plasmid library generation

Patient envelope amplicons were cloned in frame into HIV-1 molecular clone pNL4-3.Luc.R$^-$.E$^-$ (Dr. Nathaniel Landau, the NIH AIDS Reagent Program) using Gibson Assembly® Master Mix (New England Biolabs) following the manufacturer's protocol (Connor et al., 1995; He et al., 1995). An AfeI site (AGCGCT) that does not alter the amino acid sequence was introduced at nucleotide 5954 of pNL4-3.Luc.R$^-$.E$^-$ by site directed mutagenesis. A NotI site is present in the vector at the 4th codon of *nef*, the position at which the luciferase gene was inserted. The vector was linearized using AfeI and NotI. The insert was prepared by PCR amplification of the patient envelope amplicons using the following primers:

NL4-AfeI-EnvF:
(5′-TTGTTTCATGACAAA AGCGCT AGGCATCTCCTATGGCAGGAAG -3′)
NL4-NotI-EnvR:
(5′-TTTTGGCGTCTTCAGCGGCCGCGCCACCCATCTTATAGCAAAA TCCTTTC -3′).

The PCR was set up with Q5 High-Fidelity 2 × Master Mix (New England Biolabs) following the manufacturer's protocol, and run on a BioRad T100 thermal cycler (BioRad) using the following conditions: 98 °C for 30 s; 15 cycles of 98 °C for 10 s, 55 °C for 20 s, and 72 °C for 2 min; and 72 °C for 5 min. The PCR product included flanking regions of pNL4-3.Luc.R$^-$.E$^-$. The Gibson assembly reaction was performed with an insert to vector ratio of 3:1 at 50 °C for 60 min. The assembly product was diluted 1:3 and electroporated into ElectroMAX™ Stbl4™ competent cells (Thermo Fisher) in 1 mm cuvette at 1.2 kV, 25 μF, 200 Ω mA on BioRad MicroPulser™ (BioRad). Transformants were recovered in 1 mL of S.O.C. medium by shaking at 30 °C for 90 min at 225 RPM, and plated on Luria-Broth agar plates supplemented with 150 μg/mL of ampicillin (Sigma-Aldrich).

Plasmid libraries of envelope amplicons were generated by harvesting 10^6 plate-grown transformant colonies and extracting plasmid DNA using QIAGEN Plasmid Plus Mega Kit (Qiagen). Transformed stbl4 cells were plated on 150 mm ampicillin selection plates at 10^4 colonies per plate. The colonies were incubated at 30 °C for 36 h before being harvested using a scraper. The harvested bacterial colonies were centrifuged at 6000 g at 4 °C for 15 min in JA-14 rotor in a J2-HS centrifuge (Beckman). Endotoxin-free plasmid DNA libraries were prepared according to the Qiagen kit protocol. Plasmid library DNA samples for Illumina sequencing were prepared by restriction enzyme digestion using NotI and AfeI, followed by agarose gel purification of the 3 kb band corresponding to the size of the insert envelope gene.

2.3. Cell lines

HEK 293T cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal bovine serum, 2 mM L-glutamine, 1 mM sodium pyruvate, and 1x antibacterial, antimycotic solution (containing 100 U/mL penicillin, 100 μg/mL streptomycin, and 250 ng/mL amphotericin B). U87.CD4 cells were stably transfected with pBABE-CCR5-GFP or pBABE-CXCR4-GFP, constructed as previously described, and maintained in DMEM supplemented with 15% fetal bovine serum, 2 mM L-glutamine, 1 mM sodium pyruvate, 1x antibacterial, antimycotic solution, 0.2 mg/mL G418, and 1 μg/mL puromycin (Pontow and Ratner, 2001).

2.4. Virus production and passage

Replication-competent HIV-1 was produced in culture by transfecting 15 μg of the molecular clone plasmid DNA into 4 × 10^6 HEK 293T cells using *Trans*IT®-LT1 Transfection Reagent (Mirus). The viral supernatant was harvested at 48–72 h post transfection and passed through a 0.45 μm syringe filter to remove cell debris. Freshly prepared viral stocks were passaged on U87.CD4.CCR5 and U87.CD4.CXCR4 cells at low multiplicity of infection (MOI equal to or less than 0.1), in the presence of 8 μg/mL DEAE dextran, as previously described (Pontow and Ratner, 2001). U87.CD4.CCR5 and U87.CD4.CXCR4 cells were incubated with viral supernatant at

37 °C, 5% CO_2, and 95% humidity for 12 h. After removing the virus and applying fresh media, infection proceeded for another 24 h to allow for a single cycle of virus replication. The viral supernatant was harvested from the infected U87 cells, number of infectious particles quantified, and virus used in the next round of passage.

2.5. HIV-1 luciferase reporter-based titration assay

Viral stocks harvested from 293T or U87 cells were quantified for infectivity on U87.CD4.CCR5 and U87.CD4.CXCR4 cells using serial dilutions in culture media from 5 to 5^{10}-fold (Michael and Kim, 1999). The diluted viral stocks were used to infect 10^4 U87 cells in 6 replicates in a 96-well plate in the presence of 8 μg/mL of DEAE dextran. Infection was carried out by incubating the cells with dilutions of viral stocks for 12 h, removing the virus, and incubating for another 24 h in fresh media. Cells were harvested 24 h post-infection and lysed with 0.2% Triton-X 100 (Sigma-Aldrich) in PBS. The cell lysates were read for luciferase activity on Optocomp I luminometer (MGM Instruments). A result was scored as positive if the relative light unit readout was more than 2 standard deviations over the mean of mock infected control wells. The number of positive and negative wells were counted, and used to compute tissue culture infectious units using the Spearman-Karber formula.

2.6. Functional library DNA preparation

Functional libraries were prepared by PCR amplification of the envelope gene from integrated proviral DNA using genomic DNA of infected U87.CD4.CCR5 and U87.CD4.CXCR4 cells, harvested 24 h post infection using the DNeasy Blood & Tissue Kits (Qiagen) according to the manufacturer's protocol for nucleated cells. To include maximum number of quasispecies and to minimize PCR introduced founder effects, the entire genomic DNA preparation was used in multiple independent PCR runs using NL4-AfeI-EnvF and NL4-NotI-EnvR primers, and the resultant PCR products were pooled. To minimize amplification bias introduced by varying template copy number, between 400–500 ng of genomic DNA was used as PCR template for each run, resulting in approximately 6000 copies of proviral genome per reaction, after adjusting for MOI and cell number. The final pooled PCR products were purified on agarose gels and quantified on TAKE3 plates using Synergy H4 Hybrid Multi-Mode Microplate Reader (Bio-Tek).

2.7. Next generation sequencing (NGS)

DNA samples were submitted to the McDonnell Genome Institute at Washington University in St. Louis, MO for library construction and sequencing using HiSeq 2000 and/or Pacific Biosciences (Pac Bio) RS II. For amplicon and plasmid library sequencing with Illumina HiSeq2000, a minimum of 100 ng DNA per sample was used for library construction. The yield and size of input DNA was determined by a HS Qubit assay for quantitation (Invitrogen) and run on an Agilent Bioanalyzer 2100 (Agilent Technologies). Small insert dual indexed Illumina paired end libraries were constructed with the KAPA LTP Library Prep kit (KAPA Biosystems). 100 ng of genomic DNA was fragmented using the Covaris LE220 DNA Sonicator (Covaris) to a size range between 200–800 bp using the following settings: volume=50 μL, temperature=4 °C, duty factor=15%, peak incident power=450, cycle burst=200, time=130 s. The Illumina ligations were amplified in eight 50 uL reactions and were amplified with 10 PCR cycles. Libraries were fractionated on the LabChip XT using the DNA 750 chip (Perkin Elmer) collecting a 575 bp fraction with a +/− 5% covariance, followed by an AMPure XP bead purification to remove residual small fragments. Each fraction/library was assessed for concentration and size to determine molarity using the HS Qubit assay and the Agilent BioAnalyzer High Sensitivity DNA Assay. The concentration of each library fraction was verified through qPCR according to the manufacturer's protocol (Kapa Biosystems) to produce cluster counts appropriate for the Illumina HiSeq2000 platform. The libraries were pooled in equal molar ratios and loaded on 1 lane of the HiSeq2000 platform utilizing a 2×101 bp recipe according to the manufacturer's recommendations (Illumina).

Functional libraries of 3 kb in length were sequenced with HiSeq2000 as well as Pac Bio RS II. A minimum of 2 μg of functional library *env* PCR product per sample was submitted for Pac Bio library construction using SMRTBell™ Template Preparation Kit (Pacific Biosciences) following the manufacturer's protocol. Library construction input was 750 ng per library. Each sample was run on a single SMRT cell on the PacBio RS II platform using P6v2/C4 chemistry and 240 min movie length.

2.8. NGS data analysis

Illumina datasets were quality controlled by performing adaptor trimming, quality trimming, complexity screening, and length filtering using in-house codes. The amplicon and plasmid library Illumina datasets were mapped, using BWA (Li and Durbin, 2009) or Bowtie2 (Langmead and Salzberg, 2012), to sample-specific full-length reference sequences, which were predetermined by Sanger-sequencing of single clones. Pac Bio datasets of the functional libraries were quality trimmed using the SMRT Portal system to extract the high-quality Reads-of-Inserts (ROIs). The ROIs were aligned to sample-specific references using BWA-MEM. Reads that span the entire V3 region were extracted and numerated, assuming every individual read represents a single DNA molecule in the sequencing process. Rare sequences that were occurring at single-digit read level, visibly different from V3 and likely arising from mapping error and frame-shift translation were manually removed. Single Nucleotide Polymorphism (SNP) analysis was performed using GATK Haplotype Caller (https://www.broadinstitute.org/gatk/).

2.9. Phylogenetic and statistical analysis

Hierarchical clustering was performed based on Euclidean distance matrices and visualized using GENE-E (http://www.broadinstitute.org/cancer/software/GENE-E/) or iTOF (http://itol.embl.de/). V3 loop consensus sequences from two functional clusterings and their significantly different amino acid positions were computed and visualized using IceLogo (http://iomics.ugent.be/icelogoserver/index.html). The statistical significances of the differences in entropy at every nucleotide position between two samples were determined by Student's *t*-test.

2.10. Deep sequence data sets

Raw sequence files from this study will be deposited in the NCBI Short Read Archive at http://www.ncbi.nlm.nih.gov/sra under BioProject ID PRJNA317236; accession numbers: SAMN04633238-04633241.

3. Results

3.1. Evaluation of HIV-1 patient samples by deep sequencing

HIV-1 patient samples were collected from eight study subjects at two time points on ACTG A5211 Phase II study of Vicriviroc (VCV) (Table 1). Based on the original Trofile in vitro phenotypic

Table 1
Longitudinal changes of HIV-1 envelope tropism in VCV treated patients.

[a]Subject ID	Sample ID	Time (weeks)	R5-RLU[b]	X4-RLU[b]	Tropism[c]
1	1.1	0	593008	86	R5
	1.2	2	139491	185312	DM
2	2.1	0	7360	129	R5
	2.2	8	6260	2034	DM
3	3.1	0	774210	388	DM[d]
	3.2	24	223785	1178193	DM
4	4.1	0	434886	76	R5
	4.2	32	6627	117574	DM
5	5.1	0	37210	66	R5
	5.2	8	50461	78889	DM
6	6.1	0	511349	1920	DM
	6.2	8	124983	173363	DM
7	7.1	0	90578	63	R5
	7.2	8	8211	9836	DM
8	8.1	0	391164	73	R5
	8.2	2	1789	151040	DM

[a] Subject: HIV-1 patients who experienced virological failure (defined as smaller than 1 $\log_{10}$ viral load decrease at or after week 16 by protocol) in a Vicriviroc phase IIb clinical trial.
[b] RLU: Relative Light Units by luciferase-based Trofile® assay on CCR5 (R5) or CXCR4 (X4) expressing cell lines.
[c] Tropism: the ability of the virus to utilize R5 and/or X4 for viral entry.
[d] DM: dual/mixed tropism. Dual tropism: the ability to utilize both R5 and X4 for entry; mixed tropism: coexistence of viral species that can use either R5 or X4 coreceptor for entry.

assay, most study participants (all but Subjects 3 and 6) harbored only R5 viruses before VCV treatment (Time point 1 at Week 0). At the second time point, where the subject experienced protocol-defined virologic failure (less than 1 log10 decrease of HIV-1 RNA level at or after Week 16), plasma *env* amplicons demonstrated the ability to utilize CXCR4 for entry, and their tropism was designated as dual-mixed (DM). The time elapsed between the first and second time points varied among the study subjects from 2 to 32 weeks.

To understand HIV-1 *env* sequence characteristics with respect to their function, we devised a novel experimental scheme that allowed high-throughput phenotypic tropism determination followed by deep sequencing of functionally validated *env* libraries (Fig. 1A). By generating a heterogeneous library of replication competent HIV-1 virions that contained a diverse set of *env* quasispecies from patient samples, we physically separated the *env* quasispecies based on their tropism by passaging the virus on CCR5-or CXCR4-expressing U87.CD4 cells. The subset of *env* variants extracted from the proviral DNA in the infected CCR5-or CXCR4-expressing cells was termed a functional library, and their sequences were determined on both Illumina and Pac Bio platforms. To confirm that the sequence heterogeneity was not lost during the process of molecular cloning, we analyzed the original patient amplicon samples (named amplicon libraries) and the intermediate plasmid library containing one million colonies using Illumina sequencing.

A clustering analysis using single nucleotide polymorphism (SNP) information of the sixteen amplicon libraries positioned the pair of samples from each study subject on a single branch of a Euclidean-distance tree (Fig. 1B). The clustering pattern was consistent with the known sample identities, and confirmed that deep sequencing data can be used to accurately categorize HIV-1 *env* samples based on their sequence signatures.

3.2. Sequence diversity of patient amplicon libraries over time

We first analyzed the V3 coding sequence, as it is the primary determinant of coreceptor use. We found the patient samples were diverse and heterogeneous, to a similar degree as previous studies had reported, by counting the number and frequencies of different variants in the sample (AM 07:56 20-04-2016Archer et al., 2012; Tsibris et al., 2009). An in-depth population level variant analysis was performed by extracting and enumerating reads that span the entire 33 codons between the two cysteine residues at the

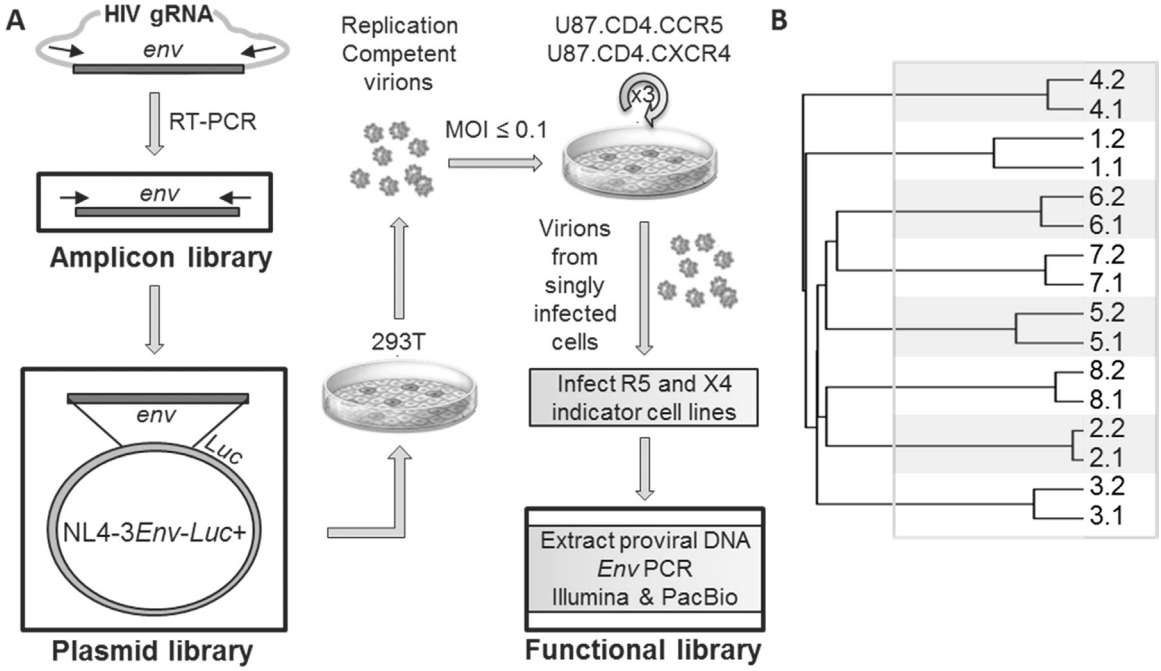
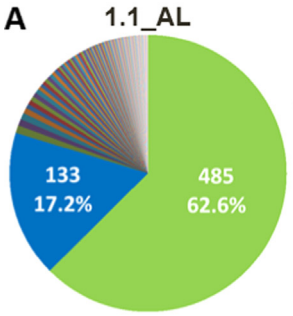
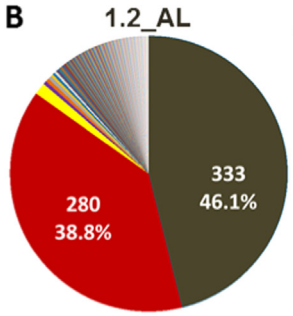


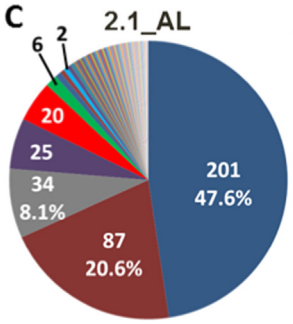
Fig. 1. (A) Schematic diagram of the experimental design. Patient derived envelope quasispecies amplicons (amplicon libraries) were cloned into replication competent NL4.3*Env-luc+* reporter vector (plasmid libraries) and expressed in 293T cells as replication competent virions that were used to infect CCR5 and CXCR4 expressing cell lines. The functionally validated *env* quasispecies (functional libraries) were prepared by polymerase-chain reaction of the proviral DNA using *env*-specific primers from the genomic DNA of the infected cells. The three sets of libraries were sequenced by Illumina HiSeq2000; functional libraries were also sequenced by Pacific Biosciences RS II. (B) SNP based clustering analysis of 16 amplicon libraries based on Illumina sequencing of the full-length *env* gene. SNPs were called with HXB2 *env* as the reference sequence using GATK. A Euclidean-distance based clustering of 16 amplicon libraries was constructed using 843 SNPs present in all samples.



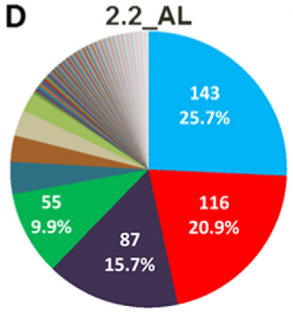
Variant #	V3 sequences	# of times	Abundance (%)	Use X4? Geno2Pheno
Sample 1.1: 93 unique V3 reads out of 775 total V3 reads				
1	CIRPNNNTRKGIHIGPGRAFYATGEIIGDIRQAHC	485	62.58	N
2	----------------G----------	133	17.16	N
3	----------------E----------	7	0.90	N
4	--G----------	6	0.77	N
5	----D----------	6	0.77	N



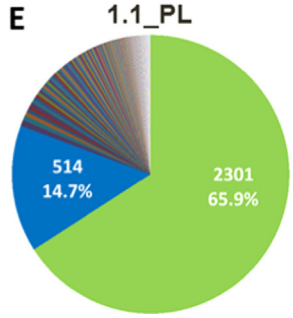
Sample 1.2: 88 unique V3 reads out of 722 total V3 reads				
1	CTRPNNNTRKGIHIGPGRAVYVAEKIIGDIRKAYC	333	46.12	Y
2	----------K----	280	38.78	Y
3	--G----------	10	1.39	Y
4	----------S----------	3	0.42	Y
5	-----S----------	3	0.42	Y
6	-------A----------	3	0.42	Y

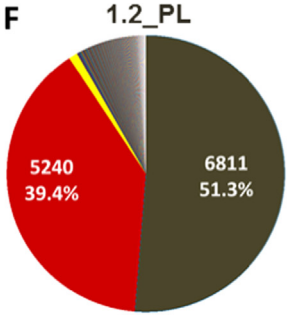


Sample 2.1 AL: 84 unique V3 reads out of 422 total V3 reads				
1	CVRPNNNTRRSINMGPGRVFHTSEIIGDIRQAHC	201	47.63	N
2	---------K----------	87	20.62	N
3	---------K----------N------	34	8.06	N
4	----S----------	25	5.92	N
5	----S----K----------N------	20	4.74	N
6	----S----------N------	6	1.42	N
7	----------N------	3	0.71	N



Sample 2.2 AL: 111 unique V3 reads out of 556 total V3 reads				
1	CVRPNNNTSKRVGIGPGMAFRATRIIGDIRQAHC	143	25.72	Y
2	----S---R-SINM---RV-HTSE---N------	116	20.86	N
3	----------A----------	87	15.65	Y
4	----S---RRSINM---RV-HTSE---N------	55	9.89	N
5	--------R-SINM---RV-HTSE----V-----	21	3.78	5.4%*
6	----S---R-SINM---RV-HTSE----------	17	3.06	N
7	----------K----------	17	3.06	Y
8	--------RRSINM---RV-HTSE---N------	13	2.34	N





G

Sample ID	Library type	Total V3 reads	Unique V3 reads
1.1	Amplicon	775	93
	Plasmid	3490	215
1.2	Amplicon	6256	236
	Plasmid	13284	329
	Clonal ctrl	31138	8

Fig. 2. Population analysis of V3 loop sequences and frequencies in amplicon and plasmid libraries. (A) Sample 1.1 amplicon library (AL), (B) Sample 1.2 AL, (C) Sample 2.1 AL, (D) Sample 2.2 AL, (E) Sample 1.1 plasmid library (PL), and (F) Sample 1.2 PL. (G) A summary table compares amplicon libraries versus plasmid libraries for Sample 1.1 and 1.2. Total V3 reads are representative of the depth of sequencing. A clonal control *env* from Sample 1.2 was prepared and sequenced in parallel. For amplicon and plasmid libraries in (A)–(F), unique V3 variants and their respective frequencies are represented as slices on a pie chart. The same color in different libraries represents a common V3 variant found in both libraries. For amplicon libraries in (A)–(D), the sequences and proportions of the most abundant V3 loop variants are listed. The coreceptor usage predicted by $\text{Geno2Pheno}_{[\text{coreceptor}]}$ is shown for every sequence. The significance level of the prediction was set at 2% and 5.75% False-Positive Rate (FPR) as the optimized cut-offs based on clinical data from the MOTIVATE study (FPR < 0.2%: X4-capable; FPR > 5.75%, X4-incapable.) *For the 5th most abundant variant in Sample 2.2, the V3 sequence is predicted to be X4-capable with a FPR of 5.4%, which falls in between the optimized cut-off values.

beginning and the end of V3. The assumption was that one read comes from one DNA molecule, which comes from one quasispecies in the infected individual. Thus, by enumerating the non-redundant reads, we captured a snapshot of the quasispecies landscape in the patient at the time of sampling (Tsibris et al., 2009). To control for PCR introduced mutations and sequence errors, we performed PCR amplification and deep sequencing of a clonal control *env* (data not shown). We detected 233 distinct non-redundant variants out of 31,989 extracted V3 reads, with the most abundant variant that was identical to the pre-determined sequence occurring 30,967 times (96.8%). None of the remaining 232 variants exceeded 0.05% of the total population. Therefore, we concluded that the error rate was low and applied a cut-off filter at 0.05% level, or at single read level to the sixteen amplicon library datasets to remove possible artifact reads.

From the sixteen samples coming from eight study subjects, we selected four samples from two subjects (Samples 1 and 2, Table 1) for the population level variant analysis (Fig. 2). We chose to focus on these two subjects because their samples demonstrated different lengths of time for coreceptor shift. Subject 1 showed high R5-tropic viral activity at baseline and comparable R5-and X4-tropic activity in only two weeks, whereas samples from Subject 2 remained predominantly R5-tropic at Week 0 and Week 8, but also showed significantly increased X4-tropism at the second time point, based on the luciferase assay readings (Table 1). Examination of the variant profile revealed distinct quasispecies compositions in these two subjects (Fig. 2A–D).

The viral population in Subject 1 before VCV treatment existed as one dominant variant that accounted for 62.6% of the total population (Fig. 2A Variant 1), a less dominant variant at 17.2% level (Fig. 2A Variant 2), and numerous rare variants that occurred in single-digit counts and less than 1% of all V3 reads. On the other hand, the second time point sample (1.2) contained 2 major variants, at 46.1% and 38.8% abundance, respectively (Fig. 2B Variant 1 and 2), followed by a minor variant at 1.4% level, and rare variants at less than 1% level. Contrary to previous studies and our expectation that a pre-existing CXCR4-using variant at Week 0 expanded under drug selection to become a major variant at Week 2, we did not detect the major variants from Sample 1.2 (Fig. 2B Variant 1 and 2) in Sample 1.1. Instead, a completely different and CXCR4-using population emerged and rapidly took over in two weeks, resulting in escape from the VCV therapy and virologic failure in Subject 1.

Different from Subject 1, Subject 2 showed dynamic expansion and contraction of variants over the 8 weeks of VCV treatment. At Week 0, the subject harbored multiple major variants, each at 47.6%, 20.6%, and 8.1% of the total, respectively (Fig. 2C Variant 1–3), all of which were predicted to use CCR5 exclusively for viral entry based on the Geno2Pheno$_{[coreceptor]}$ predictive algorithm. Some minor R5-tropic variants from Sample 2.1, present at abundances of 5.9%, 4.7% and 1.4% (Fig. 2C Variants 4–6), respectively, persisted and expanded through the treatment and were detected again at Week 8 at relative abundances of 15.7%, 20.9%, and 9.9% (Fig. 2D Variants 3, 2, and 4). The most abundant variant in Sample 2.2, accounting for 25.7% of total reads (Fig. 2D Variant 1), arose from a rare variant at 0.47% level in Sample 2.1. In contrast, the two most abundant variants present at 47.6% and 20.6% frequency in Sample 2.1 were completely eliminated from the circulation. Overall, Subject 2 largely retained its population diversity over 8 weeks, with expansion of pre-existing CXCR4-using variants and maintenance of several variants using CCR5 only in face of VCV treatment.

3.3. Evaluation of plasmid and functional library variants

To phenotypically test a large number of *env* quasispecies, we generated plasmid libraries of complete NL4.3.Luc+ molecular clones containing *env* sequences from the two samples of Subject 1. The plasmid libraries were sequenced with Illumina HiSeq to ensure that the original quasispecies diversity was retained by the cloning process. We conducted the population level variant analysis on the plasmid library sequencing datasets and discovered almost identical variant distributions, with a maximal change of major variant abundance at 3.3% of the total V3 reads (Fig. 2E compared to 2A, 2F compared to 2B, and 2G).

Functional libraries were generated by passaging plasmid libraries 1.1 and 1.2, through U87.CD4.CCR5 or U87.CD4.CXCR4 cells at low MOI (≤ 0.1) for two or three passages each consisting of a single-cycle infection. A functional library control was also generated by passage of the single *env* control in parallel. The four functional libraries obtained through this procedure included two independent passages of Sample 1.1 on CCR5-expressing cells (thus named 1.1_R5_FL1 and 1.1_R5_FL2), passage of Sample 1.2 on CXCR4-expressing cells (named 1.2_X4_FL), and a control passage of the single clone isolate from Sample 1.2 on CXCR4-expressing cells (named 1.2_Ctrl_X4). The functional libraries were sequenced by Pac Bio RSII in addition to Illumina, so that the linkage information between SNPs longer than 100 bases apart could be studied.

High-quality Reads-of-Inserts (ROIs) of the 4 functional libraries from the Pac Bio platform were generated. A total of 26,046 ROIs were obtained passed a quality filter of 99% per base accuracy, including 4667 ROIs from 1.1_R5_FL1, 8707 ROIs from 1.1_R5_FL2, 6783 ROIs from 1.2_X4_FL, and 5889 ROIs from 1.2_Ctrl_X4. The average length for ROIs is 2260 bp, with the majority of ROI at 2.9 kb, consistent with the input fragment size. All ROIs were generated from circular consensus sequencing of 10 passes or more, resulting in a low error rate per base and high quality output sequencing reads.

Analysis of variants at the population level in the four functional libraries was performed to understand the quasispecies profile across the envelope variable regions (Fig. 3). Insufficient reads from the V1 region were extracted to permit meaningful analysis, likely due to the incompatibility between our stringent read extraction criteria and the presence of large insertions and deletions in this region. Therefore, we focused on the V2 through V5 regions instead. To control for the errors introduced by PCR, virus passaging, and sequencing procedures, we analyzed the clonal control 1.2_Ctrl_X4 functional library at V2 through V5, and detected minimal amounts of deviant variants across the five regions (data not shown). Overall, in the three remaining libraries, the analysis revealed distinct quasispecies landscapes for R5 versus X4 variants. The two biological replicates, 1.1_R5_FL1 and 1.1_R5_FL2, were highly consistent in terms of the major variants and their abundances across V3–V5 regions, confirming that the passaging method was a reliable and reproducible approach to capture functionally validated quasispecies (Fig. 3, Columns 1 and 2). In R5-tropic functional libraries from 1.1, one major clone and one minor clone were found in V2, whereas V3 and V4 primarily show one major clone. The V5 regions are more diverse and consist of multiple variants each at lower abundances. In the X4-tropic functional library from 1.2, V2 through V4 consistently show two major clones, while V5 shows only one dominant clone.

Examination of the quasispecies profiles in the V3 region of the functional libraries show that they were highly similar to those of the starting amplicon libraries (Fig. 2A–B, Fig. 3). The most abundant variants in Sample 1.1 and 1.2 amplicon libraries were present in the functional libraries, suggesting the V3 variants observed in

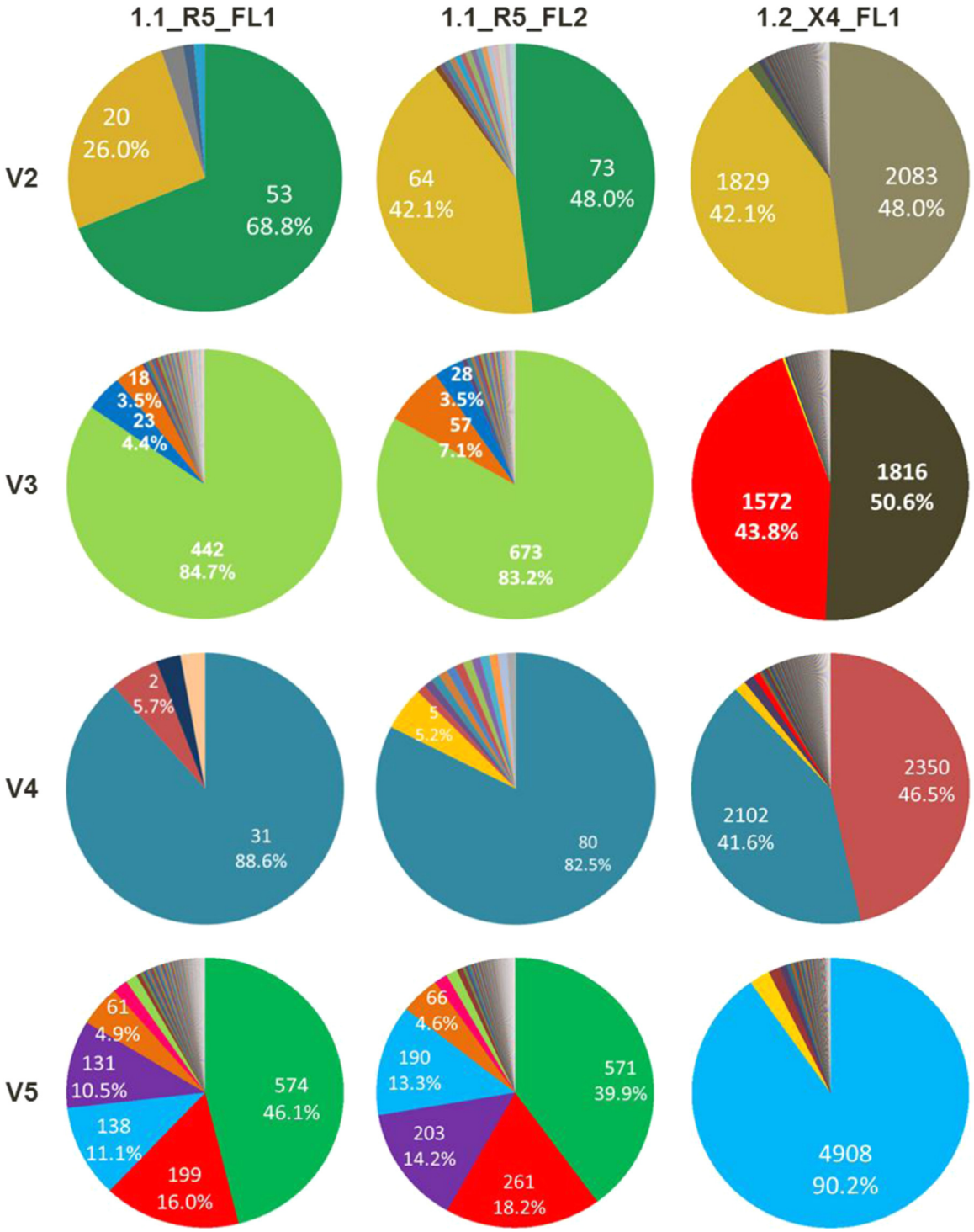


Fig. 3. Functional library population level variant analysis at *env* variable regions in Patient 1. Reads extracted from variable regions V2 (HXB2 gp160 amino acid 158–186d), V3 (297–330), V4 (385–418) and V5 (460–469) from functional libraries 1.1_R5_FL1, 1.1_R5_FL2, and 1.2_X4_FL1 are visualized in pie charts with the most abundant variants annotated. 1.1_R5_FL1 and 1.1_R5_FL2 are biological replicates prepared from Sample 1.1 passaged on U87.CD4.CCR5 cells. 1.2_X4_FL1 was prepared from Sample 1.2 passaged on U87.CD4.CXCR4 cells. Sequencing data shown were generated on the PacBio RS II system. Individual reads that span through the entire region between the given coordinates were selected, trimmed, and translated into amino acid sequences. Quasispecies are represented by non-redundant amino acid sequences and are color-coded across libraries in the same row. V3 variants are represented in same colors in Row 2 as in Fig. 2A–B, and E–F.

the amplicon libraries were part of full-length, functional *env* quasispecies. The two most abundant variants in Sample 1.1 (Fig. 2A Variants 1 and 2), as well as the three most abundant variants in Sample 1.2 (Fig. 2B Variants 1–3) were detected at comparable proportions in the functional libraries (Fig. 3, Row 2). Interestingly, a rare variant in the original amplicon library of Sample 1.1 at only 0.13% level (Fig. 2A and B) expanded through the passaging to become the third most abundant variant in 1.1_R5_FL1 at 3.5%, and the second most abundant variant in 1.1_R5_FL2 at 7.1% (Fig. 3, Row 2). This suggested that the enrichment and contraction of variants in the functional library were not stochastic events. The ability of certain variants but not the others to persist through the passaging process was directly dependent on their capacity to facilitate cell entry using CD4 and

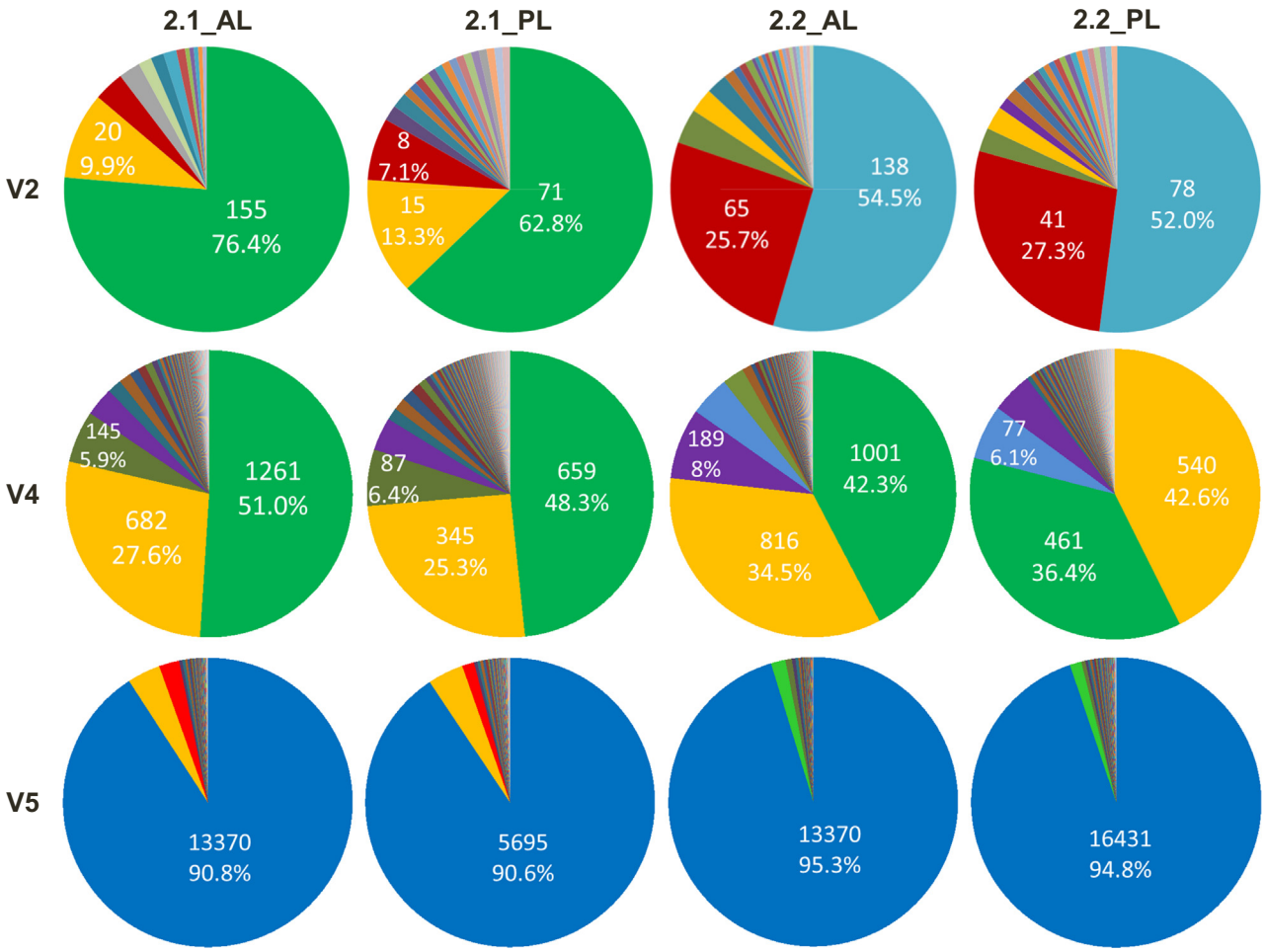


Fig. 4. Amplicon and plasmid library population level variant analysis at *env* variable regions in Patient 2. Reads extracted from variable regions V2 (HXB2 gp160 amino acid 158–186d), V4 (385–418) and V5 (460–469) from amplicon (AL) and plasmid libraries (PL) of 2.1 and 2.2 are visualized in pie charts with the most abundant variants annotated. 2.1_AL and 2.1_PL are technical replicates prepared from Sample 2.1. 2.2_AL and 2.2_PL are technical replicates prepared from Sample 2.2. Sequencing data shown were generated on HiSeq 2000 system. Individual reads that span through the entire region between the given coordinates were selected and translated into amino acid sequences. Quasispecies are represented by non-redundant amino acid sequences and are color-coded across libraries in the same row. These pie charts were analyzed together with V3 variants represented in Fig. 2C–D.

one of the two coreceptors at a given CD4 and coreceptor concentration of the reporter cell lines.

Similar variant population analyses through V2–V5 regions were performed with Amplicon (ALs) and Plasmid Libraries (PLs) for Samples 2.1 and 2.2 (Fig. 4). Comparing the ALs and PLs (Fig. 4, Column 1 versus Column 2, and Column 3 versus Column 4), it is worth noting that the major variants are mostly consistent in terms of their identity and frequency, offering additional evidence that the cloning process to generate PL from AL does not significantly alter the quasispecies representation in the community. Interestingly, in Subject 2, the V2 region showed a significant change of the variant populations (represented as change of colors in the pie chart from 2.1 to 2.2), similar to the dynamics observed in V3 (Fig. 4, Row 1, Fig. 2C–D). On the other hand, the V4 and V5 region variant distributions remained largely constant from before-treatment sample 2.1 to post-treatment sample 2.2 (Fig. 4, Rows 2 and 3). This is in contrast with Subject 1, where we observed emergence of new variants and drastic changes in the population landscape during treatment (Figs 2–4). Through population level variant analysis on two subjects, we conclude that the two individuals have undergone different viral population shifts over the course of treatment.

3.4. Characterization of V3 variants based on sequence signature shows consistency with functional tropism

To dissect the sequence signature of HIV-1 *env* isolates and its relationship with the coreceptor binding function, we generated a hierarchical cluster, based on a Euclidean-distance matrix, using 167 V3 variants from all 4 functional libraries (Fig. 5A). The sequence-based clustering separated the variants into two groups that coincided almost perfectly with their function, except for three variants from CCR5 functional libraries that were grouped with variants from CXCR4 libraries (Fig. 5A, circled in green). These three variants were predicted by the Geno2Pheno algorithm to be able to also use CXCR4 for entry, suggesting that these were rare variants before treatment that could utilize CXCR4 for entry but did not expand upon VCV treatment likely due to low fitness.

Similar clustering analysis was performed using V3 sequences from functionally validated single clones of Sample 2.2 (Fig. 5B). To cross-validate the unexpected AL deep sequencing result that shows R5-only V3 variants in Subject 2 post-treatment based on Geno2Pheno[coreceptor] prediction, we functionally tested 15 full-length *env* single clones and confirmed that approximately half of them are R5-tropic only (Fig. 2C–D). A V3 region clustering analysis showed that sequence similarities placed the 15 variants

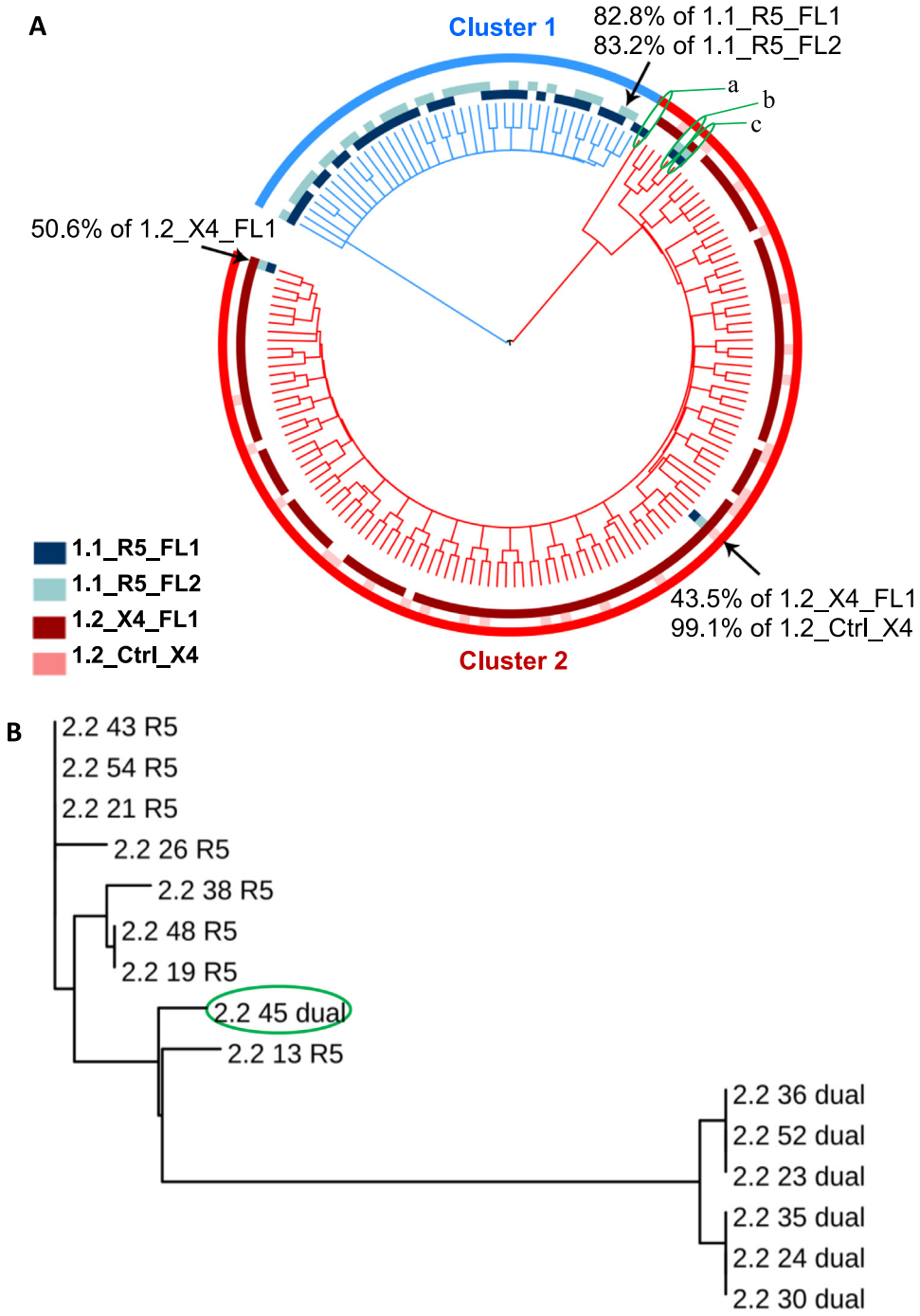


Fig. 5. (A) Hierarchical clustering analysis of 167 V3 variants from four functional libraries. (1.1_R5_FL1, 1.1_R5_FL2, 1.2_X4_FL1, and 1.2_Ctrl_X4). The presence and absence of variants is indicated using the color coded squares. Highly abundant variants are indicated by arrows. Green circles denote variants a, b, and c from 1.1_R5_FL1 and/or 1.1_R5_FL2, which were grouped in Cluster 2 and predicted to use CXCR4 by Geno2Pheno[coreceptor] (G2P). Sequences for the three circled variants are: a – CTRPNNNTRKGIHIGPGRAVYVAEKIIGNIRQAHC; b – CIRPNNNTRKGIHIGPGRAVYVAEKIIGNIRKAYC; c – CTRPNNNTRKGIHIGPGRAVYVAEKIIGNIRKAYC. (B) Hierarchical clustering analysis of 15 single clone V3 sequences from Sample 2.2 with experimentally determined differential usage of R5 and X4 coreceptors. The single clones are labeled as (sample #) (clone #) (tropism). The R5 tropic clones and dual tropic clones are clearly separately on two branches by sequence similarity, except for Clone 45 that was tested dual-tropic but clustered with other R5 tropic clones. G2P predicted both clone #45 and #13 to be R5-tropic, though #45 is only slightly over the cutoff at 2–5.74% false-positive rate (FPR) based on MOTIVATE trial. Sequences for the two most closely related but functional discordant clones are: 2.2 45 dual – CVRPNNNTRKSINMGPGRVFHTSEIIGDVRQAHC (G2P FPR 5.9%)2.2 13 R5 – CVRPSNNTRKSINMGPGRVFHASEIIGDVRQAHC (G2P FPR 9.5%).

on two branches that are mostly consistent with their function. The dual-tropic clones are grouped together and away from the R5-tropic clones, except for one clone (Clone #45) that showed more similarity to R5-tropic sequences. This is consistent with our observation in the V3 clustering analysis from Subject 1 FLs to show that V3 sequence based clustering can usually correctly predict function. However, the misplacement of clone #45 also highlights its limitations, as other regions outside V3 that contribute to coreceptor specificity are not taken into consideration in this analysis.

We generated consensus sequences of all the variants in each cluster with IceLogo2, taking into consideration their respective abundance, and identified amino acid positions differentiating cluster 1 and cluster 2 at a significance level of $P < 0.001$ (Fig. 6A–

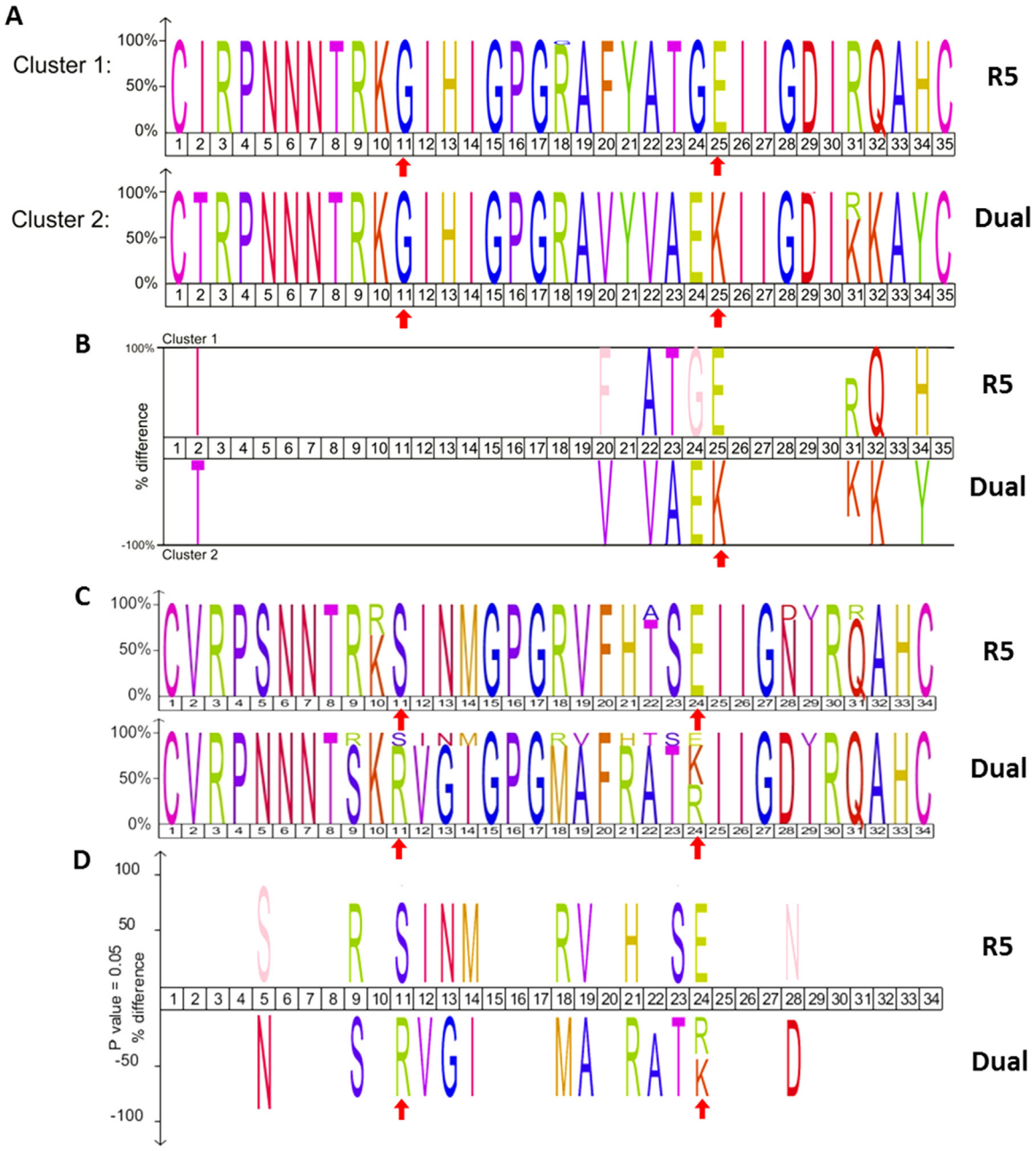


Fig. 6. (A) ConsensusV3 sequences of Cluster 1 and 2 of Patient 1 V3 variants. (B) Significantly different amino acid residues ($P < 0.001$) between the V3 consensus sequences of Cluster 1 and 2. (C) Consensus V3 sequence for 7 dual-tropic and 8 R5 tropic single clones from Patient 2 Sample 2.2. (D) Differentiating amino acid residues between the consensus V3 sequences. Due to small sample sizes, the differences are not statistically significant. Critical residues at the 11th and 24th/25th positions are highlighted with red arrows.

B). The variants in CCR5 functional libraries were highly similar to each other based on the Logo plots, and were distinct from those in CXCR4 functional libraries. Cluster 1 showed a high degree of conservation, which was reflected by a major variant that accounted for over 80% of the population. Cluster 2 showed a clear divergence at Position 31, which was representative of two major variants at 43.5% and 50.2%, respectively (Fig. 5A). Altogether, we identified nine positions (Positions 2, 20, 22, 23, 24, 25, 31, 32, and 34) in the V3 region with significantly different amino acid residues, that were likely responsible for the differential functionality to engage CCR5 versus CXCR4 as entry coreceptor (Fig. 5B). Consistent with our observation, these positions and amino acid residues have been previously reported in HIV-1 Subtype B and C patient samples to correlate with CCR5 or dual-tropism (Masso and Vaisman, 2010; Rodriguez et al., 2015). In addition, at the 25th position in Cluster 1 was a negatively charged glutamic acid, whereas occupying the same position in Cluster 2 was a positively charged lysine, which was indicative of CXCR4-usage by the 11/25 Rule (Jensen and van't Wout, 2003).

A similar analysis was performed for the two functional groups from Sample 2.2 and the findings were consistent with those from Subject 1 (Fig. 6C–D). As compared to Subject 1, we observed, in Subject 2, diversity at more positions (Positions 10, 22, 28, 29, and 31 in R5-tropic group, and Positions 9, 11–14, 18, 19, 21–24, and 29 in dual-tropic group). This result is likely due to the small sample sizes (8 for R5-tropic group and 7 for dual-tropic group), and also the diverse V3 landscape in Subject 2, as previously mentioned (Fig. 6C, Fig. 2D). Residues that differ between the two groups were identified although they failed to reach statistical significance (Fig. 6D). Residue changes at Positions 11 and 24 (instead

of Position 25, due to a deletion in the V3 sequences from Sample 2.2 compared to those from Sample 1.2) were also noted, in particular the charge flip from glutamic acid to arginine or lysine at Position 24. Through the clustering and consensus analysis from 2 subjects, we demonstrated that categorizing HIV-1 V3 variants by sequence similarity created clusters that were mostly consistent with their function. Moreover, these results confirmed that the V3 region is a critical modality for coreceptor tropism determination.

3.5. *Important regions outside V3 that correlate with envelope tropism*

We analyzed the full *env* gene to examine other determinants of tropism beyond the V3 region using Shannon's entropy approach and SNP analysis for Subject 1 and 2 datasets (Figs. 7-8). Shannon's entropy analysis examines cross-sectional diversity at a given position, with higher entropy values responding to more diversity. We utilized entropy analysis to compare the entropy values at each nucleotide position between pre-treatment and post-treatment amplicon libraries of Subjects 1 and 2, to explore regions that were undergoing an increase or decrease in diversity due to VCV selection.

Overall, in Subject 1, Sample 1.1 had higher entropy across the length of the *env* gene than Sample 1.2, as demonstrated by 120 positions in Sample 1.1 and only 51 positions in Sample 1.2 that have statistically above-background entropy values when compared with the clonal control dataset (Fig. 7A, $p < 0.0001$). This suggested that the viral quasispecies population before VCV treatment was more heterogeneous than the quasispecies present after treatment. The treatment likely selected for a dominant drug-resistant variant, or a group of closely related resistant variants, thus, driving down the overall diversity in the population. The decrease in entropy from Samples 1.1 to 1.2 occurred along the length of *env*, as represented by the entropy difference map (Fig. 7B). However, specific locations showed an increase in entropy, in particular at HR1/HR2 regions and the cytoplasmic domain of gp41, suggesting diversifying selection at those loci (Table 2).

In Subject 2, Samples 2.1 and 2.2 showed less difference in entropy than the samples from Subject 1 (Fig. 8). We identified 98 positions in Sample 2.1 and 79 positions in Sample 2.2 that have entropy values significantly over zero (Fig. 8A, $p < 0.0001$). In contrast to the reduction in entropy over treatment in Subject 1, the entropy values showed an overall increase over the course of treatment in Subject 2, suggesting diversifying selection on the quasispecies population in Subject 2 (Fig. 8B). This is consistent with our previous observation that Subject 2 had more diverse variant landscapes than Subject 1 (Fig. 2A–D). Five positions between Sample 2.1 and 2.2 experienced significant changes in entropy, all of these positions are located in gp120 (Table 2).

SNP analysis on the four functional library datasets using GATK program identified 106 SNPs in the entire *env* gene with reference to the control single clone sequence from Sample 1.2. Clustering of

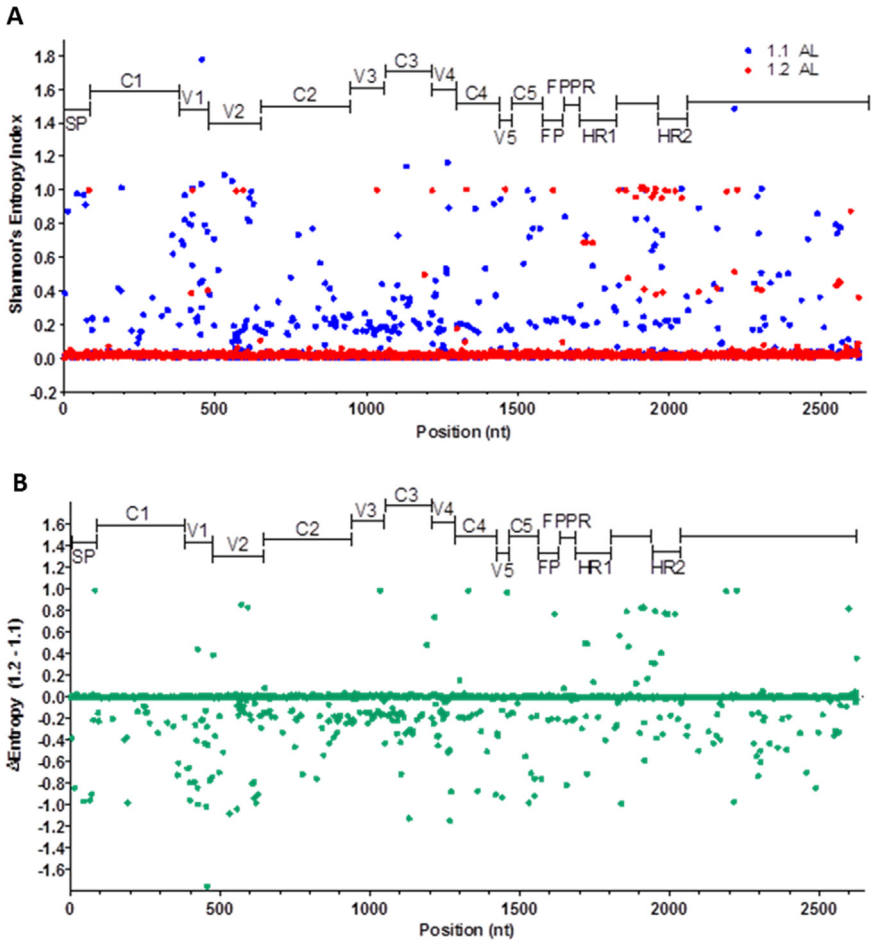


Fig. 7. Shannon's entropy analysis of amplicon libraries (AL) from Sample 1.1 and 1.2. (A) Absolute entropy values at each nucleotide position along the *env* genes for 1.1 AL and 1.2 AL. A total of 120 positions in Sample 1.1 (blue) and 51 positions in Sample 1.2 (red) have entropy values significantly over background entropy values in the 1.2 clonal control dataset (not shown), by one-tailed Student's *t*-test, $p < 0.0001$. (B) Entropy difference at each nucleotide position by subtracting the values of 1.1 AL from that of 1.2 AL. At 17 positions (shown in Table 2), the entropy values in Sample 1.1 differ significantly from those in Sample 1.2, by two-tailed Student's *t*-test, $p < 0.0001$.

Table 2
Positions in *env* that correlate with coreceptor usage.

	AA[a]	SNP[b]	ΔEnt[c]	Pt[d]	Significance	Ref
SP	8		0.59	2		
	24		−0.96	1		
	26		−0.90	1		
	29		0.99	1		
C1	86	L/M		1	gp120-gp41	Pancera et al. (2014)
	87	E/G/V/A		1, 2	gp120-gp41; N88 glycosite	
	99	N/S		2		
V1	134a		−0.96	1		
V2	168	K/R		2	gp120-gp120	Pancera et al. (2014)
	170		−0.92	2		
	178	T/K		2		
C2	229	K/R		2	C228-C239; N230 glycosite	Leonard et al. (1990)
	240		−0.72	1	C228-C239; N241 glycosite	
	254	V/I		2		
	256		−0.76	1		
V3	295		0.36	2	N295 glycosite; DC-SIGN bs	Hong et al. (2007)
	300	S/N		2	Coreceptor bs; N301 glycosite	Korber and Gnanakaran (2009)
	309		0.16	2	R5/X4 specific site	
	325	D/N		1, 2	Coreceptor bs	
	327		0.99	1		
C3	344	Q/R		2		
	351	E/K		1		
	354	K/E/R		2		
V4	412	E/D		1		
C4	421	R/K		1	Coreceptor bs	Korber and Gnanakaran (2009)
	424	V/I		1	R5 × 4 specific site	
	425	N/K		1	CD4 main/side chain bs	Zhou et al. (2010)
	432	K/T		1	CD4 main chain bs	
	440	N/R/S		1, 2	R5 × 4 specific site	Korber and Gnanakaran (2009)
	442		-0.87	1	Coreceptor bs	
	444	S/R		1	C378-C445	Leonard et al. (1990)
V5	465d	R/K		1		
	467		0.42	2		
C5	476	K/R		1, 2	CD4 side chain bs	Zhou et al. (2010)
	490	Q/K		1	gp120-gp41	Pancera et al. (2014)
	496	L/V	-0.71	1		
	500	K/M	-0.92	1		
	502	K/R		2		
	503		−0.76	1		
FP	514a	M/T/I		1		
	518	V/M		1		
	525	T/A		1	gp120-gp41	Pancera et al. (2014)
FPPR	534	A/S		1	gp120-gp41, gp41-gp41	
	535		−0.82	1	gp41-gp41	
	541	T/A		1	gp120-gp41, gp41-gp41	
	543	R/Q		1	gp120-gp41	
HR1	547	G/V		2	gp41-gp41	
	556	M/L		2		
	596		−0.99	1	gp120-gp41	Pancera et al. (2014)
	621	E/D		1		
HR2	633		0.31	1	gp120-gp41	Pancera et al. (2014)
CD	724		0.98	1		
	749		0.99	1		
	802		−0.50	1		

A total of 54 residues that correlate with coreceptor usage are identified from 2 patients (Patient 1: 38; Patient 2: 20; both: 4).
SP: signal peptide; FP: fusion peptide; bs: binding site.

[a] AA: amino acid positions are numbered based on HXB2.

[b] SNP: Positions showing > 50% switch from reference to alternative allele in SNP analysis are annotated as ref/alt/alt/alt. Reference alleles are based on 1.2 and 2.2 single clone control sequences for Patient 1 and 2 respectively.

[c] ΔEntropy = post treatment entropy − before treatment entropy. Positive value: increase in entropy/diversity; negative value: decrease in entropy. Changes in entropy are statistically significant, with p values < 0.0001 by two-tailed t-test.

[d] Pt. indicates the patient number of the datasets from which a particular residue is identified by SNP and/or entropy analysis.

the four functional library datasets based on their SNP frequency patterns reflects the functional categorization of the four libraries, with 1.1_R5_FL1 and 1.1_R5_FL2 on a single branch, and away from 1.2_X4_FL and 1.2_Ctrl_X4 on a different branch (Fig. 9A). To identify SNPs that could differentiate CCR5 versus CXCR4 coreceptor usage, we focused on 23 non-synonymous loci (annotated in the red boxes) that were completely different in CCR5-using groups than in CXCR4-using groups, including 22 loci outside the V3 region. Among the 23 positions identified, 4 positions (N325, K421, I424, and S440) were previously known to strongly affect coreceptor binding and specificity, and 3 positions (N425, K432, and R476) were reported to be CD4 binding residues (Kwong et al., 1998; Korber and Gnanakaran, 2009; Zhou et al., 2010). Sixteen positions identified through the SNP analysis have not previously been identified to have a function in receptor or coreceptor binding, although many in this unknown category (L86, N87, K490, V496, K500, A525, S534, A541, and Q543) were found to be involved with gp120 – gp41 interaction (Table 2, Fig. 10) (Korber and Gnanakaran, 2009; Zhou et al., 2010; Pancera et al., 2014; Hong et al., 2007; Leonard et al., 1990). This finding suggests a role of gp120-gp41 interaction in coreceptor specificity, potentially by affecting trimer stability, entry dynamics and thus fusion efficiency following coreceptor binding. A recently published study also identified residues 432 and 500 for their importance in trimer stability, highlighting the possibility that these interaction surface

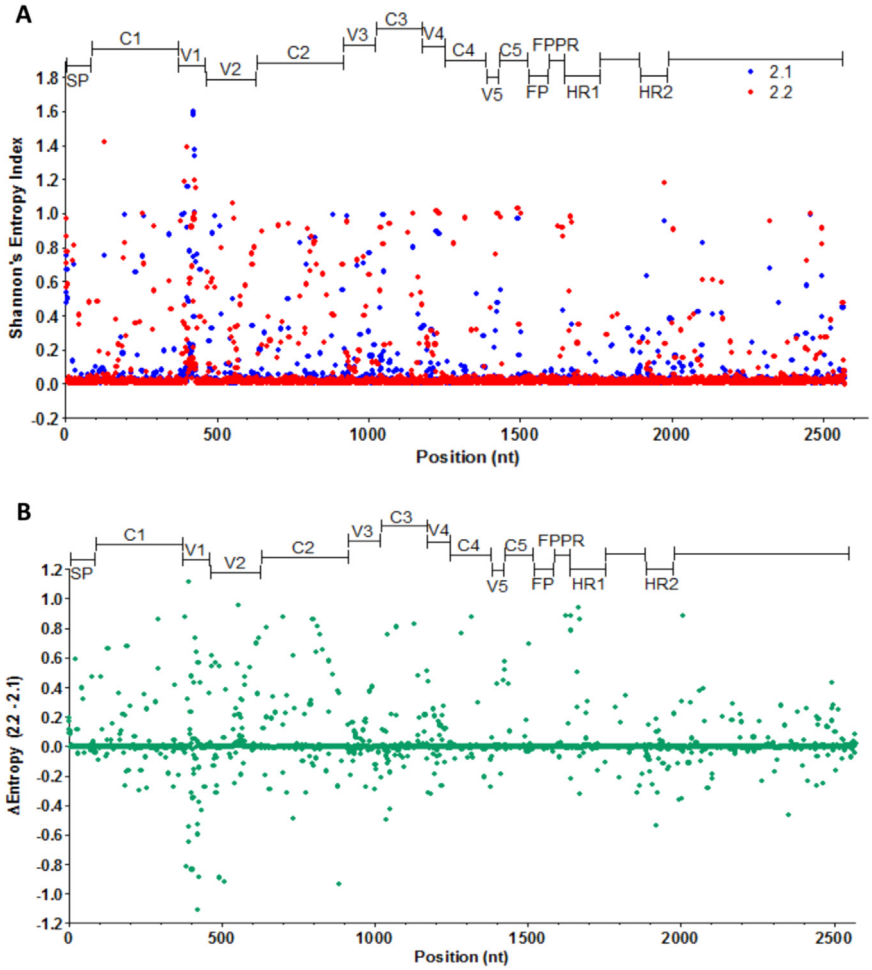


Fig. 8. Shannon's entropy analysis of Sample 2.1 and 2.2. (A) Absolute entropy values at each nucleotide position along the *env* genes for 2.1 and 2.2. A total of 98 positions in Sample 2.1 (blue) and 79 positions in Sample 1.2 (red) have entropy values significantly over zero, by one-tailed Student's *t*-test, $p < 0.0001$. (B) Entropy difference at each nucleotide position by subtracting the values of 2.1 from that of 2.2. At 5 positions (shown in Table 2), the entropy values in Sample 1.1 differ significantly from those in Sample 1.2, by two-tailed Student's *t*-test, $p < 0.0001$.

residues might play a functional role in the overall envelope trimer structure (Guenaga et al., 2015).

Similar SNP analysis was performed with Subject 2 amplicon and plasmid library datasets, based on the single clone analysis results that approximately 50% of post-treatment variants in Sample 2.2 can utilize CXCR4 (Fig. 9B, Fig. 5B). To identify residues that associate with R5-versus X4-usage, we focused on 15 non-synonymous loci (annotated in the red boxes) that showed at least 50% difference in reference allele frequency between pre-treatment and post-treatment samples. Taken together with the 5 positions identified in the entropy analysis, the total 20 loci identified include 4 previously reported coreceptor binding specific sites (N300, I309, N325, and S440), 1 in the CD4 binding site (R476), and 3 positions within or adjacent to glycosylation sites (N229, N295, and N300). Thirteen positions were not previously reported to have a function in coreceptor binding (Table 2). Interestingly, 4 loci out of the 13 are located on an interaction surface between gp120-gp41, gp41-gp41, or gp120-gp120 and could potentially affect envelope trimer stability.

Comparing these residues that associate with differential coreceptor use from Subjects 1 and 2, we identified 4 loci that are common to both subjects (V87, N325, S440, and R476). Among the 4 common residues, N325 and S440 were previously reported to be coreceptor binding specific sites inside and outside V3 respectively, R476 is within the CD4 site chain binding site, and V87 is a gp120-gp41 contact site within one subunit of the trimer. Besides the residues in common between the two subjects, many other loci from the two subjects fall into similar functional domains of CD4 interaction, coreceptor binding, and interaction between subunits in the timer, even though they are not at the exact same position. This evidence strongly suggests that there are common pathways that affect coreceptor shift. Besides coreceptor binding and CD4 binding pathways that have been extensively studied, residues that play a role in envelope structural stability potentially influence coreceptor usage.

For visualization of the critical regions that correlated with coreceptor specificity, annotation was performed on the ligand-free native envelope gp140 trimer crystal structure (PDB ID: 4MJZ) (Fig. 10). A group of residues (colored red) that were previously reported to be coreceptor-specific is located near the envelope-coreceptor binding interface and likely to affect binding via direct contact with extracellular regions of the coreceptors. A second group of residues (colored magenta) located near the CD4 binding site is likely to affect coreceptor usage via interaction with CD4 and subsequent induction of a conformational change into an open trimer. A third group (colored blue and green) that is located at contact surfaces between gp120-gp41, gp41-gp41 and gp120-gp120 is likely to affect tropism via more global mechanisms, such as alteration of the overall configuration of envelope trimer or its dynamic conformational change upon CD4 and/or coreceptor engagement.

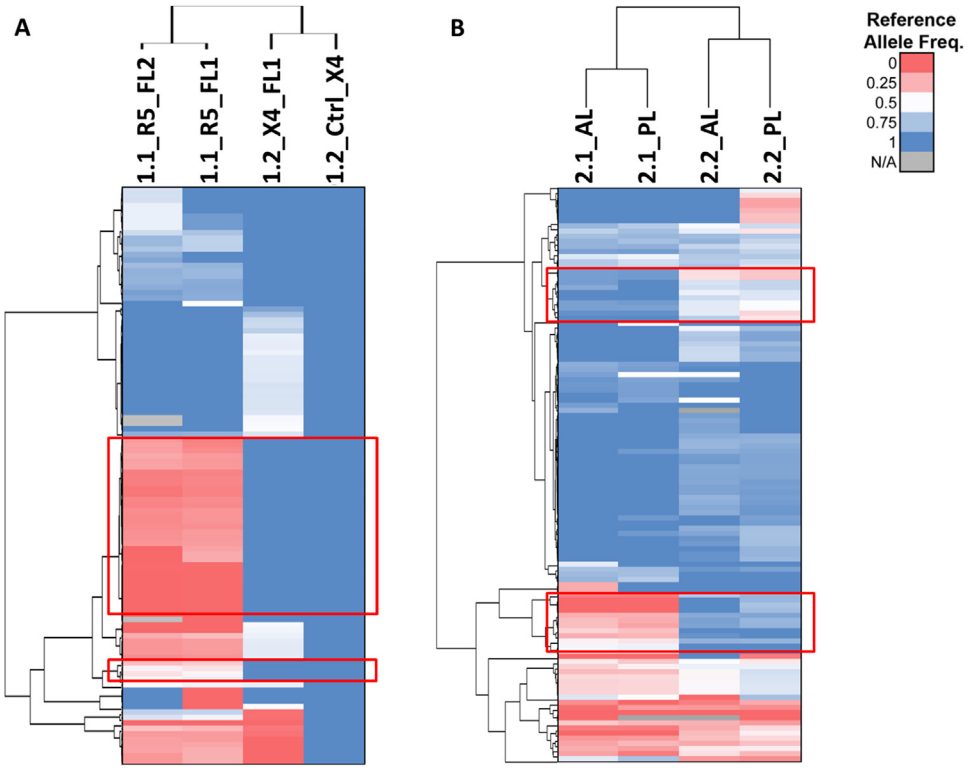


Fig. 9. Clustering analysis based on SNP loci allele frequency of full length *env* datasets from Patient 1 and 2. (A) Four functional libraries (FL) from Sample 1.1 and 1.2 were clustered by overall similarity between the four samples based on 106 SNP loci allele frequencies. A single clone control (1.2_Ctrl_X4) sequence was used as reference. A total of 37 differentiating SNPs are found in the regions outline by the red boxes (residues showing over 75% shift from reference to alternative alleles), among which 24 SNPs at 23 codons lead to non-synonymous amino acid substitution (shown in Table 2). Only 1 out of the 24 SNPs falls within the V3 loop region. (B) Amplicon libraries (ALs) and plasmid libraries (PLs) from Sample 2.1 and 2.2 were clustered by similarity between the four datasets based on 112 SNP loci allele frequencies. A single clone sequence from 2.2 was used as reference. A total of 22 differentiating SNPs are identified in the regions outlined by the red boxes (residues showing over 50% shift from reference to alternative alleles), among which 16 SNPs lead to non-synonymous amino acid substitutions (shown in Table 2). Two out of the 16 SNPs fall within the V3 region.

4. Discussion

In this study, we developed a comprehensive phenotypic assay platform in combination with NGS technologies to analyze HIV-1 quasispecies diversity and tropism before and after emergence of drug resistance in vivo. This novel approach employed a tissue-culture based library passaging method and allowed us to study in unprecedented depth and breadth the quasispecies population and their structure-function relationship by generating sequences from phenotypically validated *env* variants. The output data provided new insights into important features of HIV-1 env that are implicated in coreceptor interaction, specifically the residues that are located at the trimer interaction surface and could potentially affect envelope spike stability and conformational dynamics. These findings shed light on new drug targets for HIV-1 entry inhibitor development. In addition, the experimental approach can potentially be adapted for the study of drug resistance development in order to identify new therapeutic agents, especially for development of unusual combinations of drug resistant mutations that requires phenotypic characterization of new and complex genotypes.

Quality control of the sequencing data is extremely important in NGS-based studies to ensure the accuracy and relevance of the final conclusion. In our study, we performed all the experimental procedures including PCR, cloning, infection, passaging, Illumina and Pac Bio sequencing, reads filtering, and the final analysis with a clonal control of known nucleotide sequence in parallel. Therefore, we could calculate the cumulative error rate of PCR and cloning errors, passaging artifacts, and sequencing errors at every step of the process by analyzing the single clone control dataset and apply an appropriate filter level based on the control error rate. Notably, the combined error rate from PCR-induced substitution and sequencing in our experience were markedly lower than previously reported, likely reflecting improved fidelity ($>100\times$ of *Taq*) of the PCR polymerase and a newer generation sequencing method (Tsibris et al., 2009).

A potential source of error that we could not control for using the single clone method is strand-switching (recombination) during the PCR, transfection and infection process. Salazar-Gonzalez and coworkers demonstrated the extent of bulk PCR induced sample skewing including nucleotide substitutions and recombination by comparing the single clone sequences from bulk PCR to those obtained through Single-Genome Amplification approach (Salazar-Gonzalez et al., 2008). We estimated the effects of RT-PCR and second round PCR on amplicon libraries based on previously published information on patient samples that came from the same cohort as ours. The recombined clones resulted from PCR accounted for less than 0.15% of the total population, and did not significantly affect the proportions of major variants as demonstrated previously (Tsibris et al., 2009). For subsequent PCR-based amplifications to generate the plasmid and functional libraries, we used a high-fidelity high-processivity polymerase, low cycle number (15 cycles), and high primer concentration to suppress the level of recombination as described by Smyth and coworkers (Smyth et al., 2010). In order to minimize recombination events and production of genotype-phenotype mismatched virions from cells infected by more than one viral particle over the infection process, we performed the passages using an MOI of 0.1 or less. By Poisson distribution, through one passaging process at MOI of 0.1, less than 5% of virions come from multiply infected cells; whereas after three rounds of passaging, 0.02% of virions results from multiply infected cells that may undergo

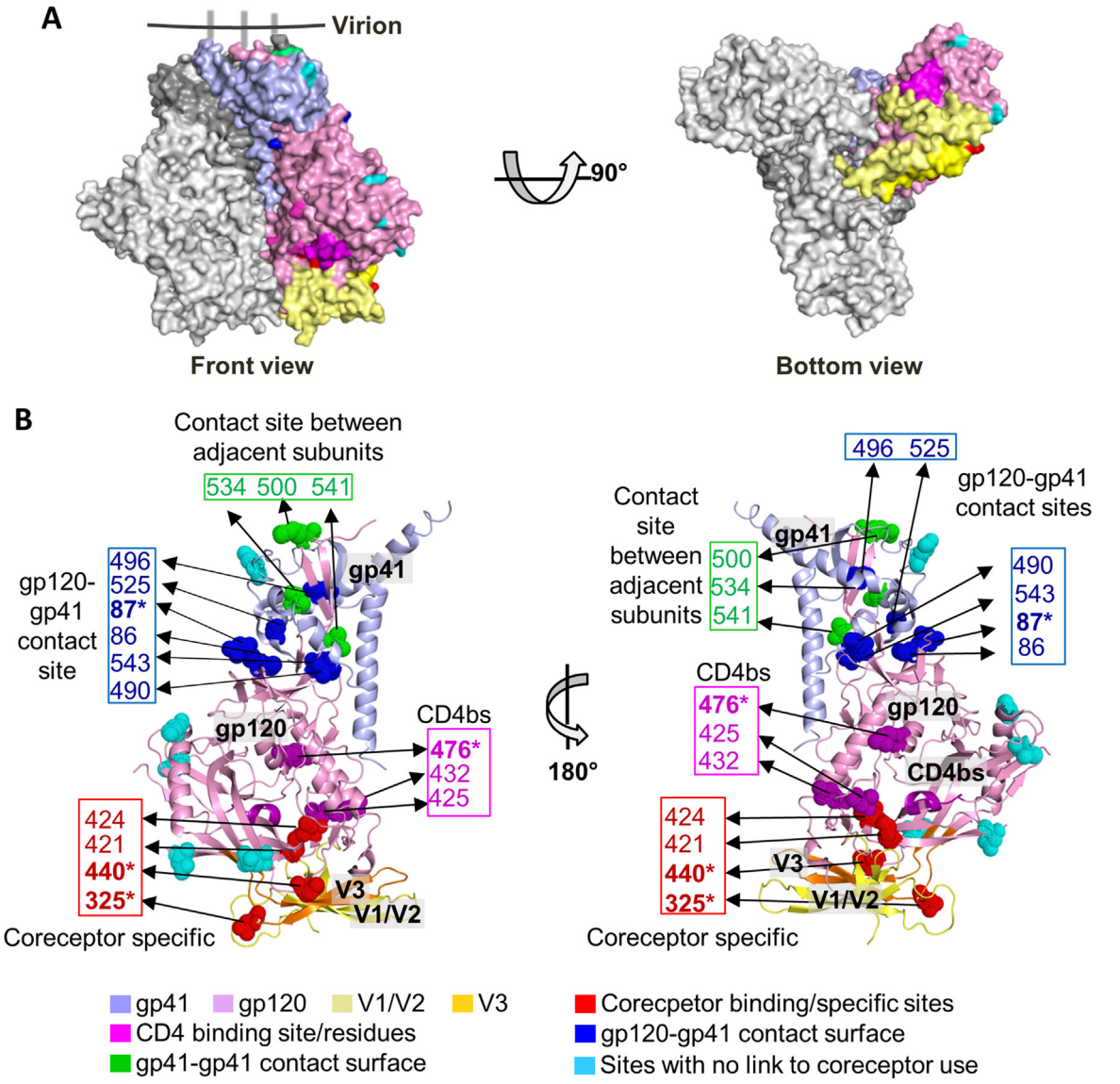


Fig. 10. Residues within and outside the V3 loop differentiate R5-versus X4-usage. (A) Representation on the gp140 SOSIP ligand-free trimer structure (PDB ID: 4ZMJ) of 16 loci differentiating CCR5 from CXCR4 utilization. The front view shows a trimeric spike with gp41 on the top and gp120 at the bottom. The bottom view shows the perspective looking up from the host cell membrane. The subunits are shown in surface model, with gp120 in light grey and gp41 in dark grey in two of the three subunits. The third subunit is colored by region as the following: light blue, gp41 ectodomain; pink, gp120; pale-yellow, V1/V2 loop; orange, V3 loop; and magenta, CD4 binding site (CD4bs). The differentiating loci are color-coded based on their previously reported known functional status. Red: coreceptor binding or specific site; magenta: CD4 contact residues; blue: intra-subunit gp120-gp41 interaction site; green: inter-subunit gp120-gp41 and gp41-gp41 interaction site; teal: no previously reported function in viral entry. (B) Enlarged front and back views of the 16 differentiating residues on a cartoon representation of gp140 subunit. Residues are annotated and grouped based on their functional status and color coded in the same manner as in (B). Four residues identified in both Patient 1 and 2 (87, 325, 440, and 476) are shown in bold with asterisk (*).

recombination during reverse-transcription, or carry mismatched genomic RNA compared to the envelope protein on the surface. In addition, we also utilized the amplicon and plasmid library dataset as baseline for the functional library variant analysis, and removed 55 out of 220 V3 variants that did not exist in the first two libraries, therefore eliminating the majority of recombination errors and passaging artifacts. Biological replicates of Sample1.1 on CCR5-expressing cell lines served as another layer of control and confirmed that our approach is robust and reproducible. Finally, genotypic and phenotypic tropism analysis of 97 single clones from Subject 1 and 2 amplicon libraries confirmed the sequences and tropism of all the dominant V3 variants identified by deep sequencing, thus demonstrating the accuracy of sequence and tropism determination by this new integrated approach (data not shown).

The quasispecies landscape at the two snapshots through time in two study subjects alluded to different evolutionary pathways. Through analysis of V3 variants in amplicon libraries from Subject 1, we were unable to detect a minor variant at Week 0 that expanded into a major variant at Week 2 despite multiple rounds of sequencing and extensive depths of coverage at 600 × to 6000 ×. If a minor CXCR4-tropic variant was present at very low frequency (less than 3.2% of the total population) in the circulation, it would not be detected by the current approach. Therefore, a plausible origin of the X4-using variants may be from one or a few very low frequency pre-treatment variants. Consistent with the "fitness valley" model of coreceptor shift, we detected few intermediate variants from CCR5-only to CXCR4-using genotypes, as they are thought to be less fit and are quickly selected out (Poon et al., 2012; da Silva and Wyatt, 2014). On the contrary, Subject 2 demonstrated a typical dynamic quasispecies population shift under the selection pressure from the CCR5 antagonist, during which rare pre-existing CXCR4-using clones expanded and

susceptible clones declined. Interestingly, several major variants predicted to be R5-tropic only in Sample 2.1 had unexpectedly persisted and expanded over treatment, perhaps by acquiring resistance through an alternative mechanism, such as binding to the CCR5-VCV complex for entry (Moore and Kuritzkes, 2009).

One limitation of this study is the sampling method. We were unable to access quasispecies residing in lymph nodes and latently infected cells, as the patient samples were collected and prepared from free circulating virions in the plasma. Since HIV-1 quasispecies at these sanctuary sites are often sheltered from antiretroviral therapy, or are exposed to a lower dose and sub-optimal combination of antiretroviral drugs, the virus may continue to replicate at low levels and develop resistance progressively in a step-wise fashion, even when the treatment appears fully suppressive (Moreno-Gamez et al., 2015; Rothenberger et al., 2015). Some of the viral quasispecies that eventually escape treatment and migrate to the blood are predicted to harbor multiple mutations, consistent with our observations. A recent study also shows that proviral populations in peripheral blood mononuclear cells are different from those in lymphoid organs and in latently infected memory cells; the variants in circulation are often defective or less fit (Cohn et al., 2015). Since we are just beginning to understand these tissue-localized HIV quasispecies, more studies are needed to examine the evolutionary process in local lymphoid tissues and latently infected reservoirs to elucidate how they contribute to coreceptor shift and resistance development.

Another caveat with the analysis to identify important residues for coreceptor shift is a small sample size. We were examining two subjects from a clinical trial. While we found common pathways and trends that would affect coreceptor usage, due to the small sample size we cannot completely eliminate founder effect in this analysis. Approximately 40% of all the residues (22 loci out of 54 total loci from two patients combined) have no previously reported functional connection to coreceptor usage or subunit interaction. Many of these could be passenger mutations that occur during the viral evolutionary process in these individuals, and do not have a role in drug resistance. While it is reassuring to see common residues identified in both patients and more residues that fall into similar categories, future studies are needed to examine more patients and to experimentally validate these residues in vitro to establish a clear functional connection to coreceptor specificity.

Overall, this study employed a novel experimental approach to characterize phenotypically validated envelope sequences with confidence, by coupling a phenotypic passaging assay with deep sequencing analysis. We identified three regions outside V3 that associate with coreceptor shift in the context of VCV selection, including the C4 region, CD4 binding site, and the interaction surface between gp120-gp41 and between gp41 subunits. A majority of the residues identified in the first two regions are consistent with and confirm previously published data on coreceptor specificity and viral entry, while a new subset of mutations at the gp120-gp41 interaction site alludes to interesting new hypotheses for the mechanism of coreceptor shift, even though it is based on the observation on a single patient before and after CCR5 antagonist treatment. One could speculate and investigate into possibilities such as fusion efficiency enhancement or global envelope quaternary structure modulations through the gp120-gp41 interaction to compensate for weak CXCR4 binding and confer X4-tropism.

Last but not least, this experimental approach would potentially be applicable for drug resistant mutation discovery and research in the context of other therapeutic agents. This study serves as a proof of concept in the context of tropism testing and CCR5 antagonist treatment. By coupling phenotypic and phenotypic testing, this integrative approach harnesses the power of deep sequencing analysis to understand the process of drug resistant development over the viral quasispecies evolution process.

Funding source

This work was supported by National Institutes of Health grant R01 AI106361 to L.R., UM1 AI068636 and R37 AI055357 to D.R.K.

Acknowledgments

We thank Dr H. Huang for assistance with PyMOL, Dr J. Fay for guidance in dN/dS analysis, Dr K. Wylie and Dr J. Gu for assistance with NGS data analysis, Drs. Kyei, Rauch, McCulley, and Leibel for intellectual discussions, Dr D. Wang and Dr. G. Zhao for diversity analysis, and Ms N. Campbell and Mr J. Zhang for technical assistance. We also thank Dr. Roy M. Gulick and members of the ACTG A5211 protocol team, as well the participating ACTG Clinical Research Sites, for providing the plasma samples used in this study.

References

Archer, J., Weber, J., Henry, K., Winner, D., Gibson, R., Lee, L., Paxinos, E., Arts, E.J., Robertson, D.L., Mimms, L., Quinones-Mateu, M.E., 2012. Use of four next-generation sequencing platforms to determine HIV-1 coreceptor tropism. PLoS One 7, e49602.

Berger, E.A., 1997. HIV entry and tropism: the chemokine receptor connection. AIDS 11 (Suppl. A), S3-16.

Cashin, K., Sterjovski, J., Harvey, K.L., Ramsland, P.A., Churchill, M.J., Gorry, P.R., 2014. Covariance of charged amino acids at positions 322 and 440 of HIV-1 Env contributes to coreceptor specificity of subtype B viruses, and can be used to improve the performance of V3 sequence-based coreceptor usage prediction algorithms. PLoS One 9, e109771.

Cohn, L.B., Silva, I.T., Oliveira, T.Y., Rosales, R.A., Parrish, E.H., Learn, G.H., Hahn, B.H., Czartoski, J.L., McElrath, M.J., Lehmann, C., Klein, F., Caskey, M., Walker, B.D., Siliciano, J.D., Siliciano, R.F., Jankovic, M., Nussenzweig, M.C., 2015. HIV-1 integration landscape during latent and active infection. Cell 160, 420–432.

Connor, R.I., Chen, B.K., Choe, S., Landau, N.R., 1995. Vpr is required for efficient replication of human immunodeficiency virus type-1 in mononuclear phagocytes. Virology 206, 935–944.

Connor, R.I., Sheridan, K.E., Ceradini, D., Choe, S., Landau, N.R., 1997. Change in coreceptor use correlates with disease progression in HIV-1–infected individuals. J. Exp. Med. 185, 621–628.

da Silva, J., Wyatt, S.K., 2014. Fitness valleys constrain HIV-1's adaptation to its secondary chemokine coreceptor. J. Evol. Biol. 27, 604–615.

Do Kwon, Y., Pancera, M., Acharya, P., Georgiev, I.S., Crooks, E.T., Gorman, J., Joyce, M.G., Guttman, M., Ma, X., Narpala, S., Soto, C., Terry, D.S., Yang, Y., Zhou, T., Ahlsen, G., Bailer, R.T., Chambers, M., Chuang, G.Y., Doria-Rose, N.A., Druz, A., Hallen, M.A., Harned, A., Kirys, T., Louder, M.K., O'Dell, S., Ofek, G., Osawa, K., Prabhakaran, M., Sastry, M., Stewart-Jones, G.B., Stuckey, J., Thomas, P.V., Tittley, T., Williams, C., Zhang, B., Zhao, H., Zhou, Z., Donald, B.R., Lee, L.K., Zolla-Pazner, S., Baxa, U., Schon, A., Freire, E., Shapiro, L., Lee, K.K., Arthos, J., Munro, J.B., Blanchard, S.C., Mothes, W., Binley, J.M., et al., 2015. Crystal structure, conformational fixation and entry-related interactions of mature ligand-free HIV-1 Env. Nat. Struct. Mol. Biol. 22, 522–531.

Dobrowsky, T.M., Rabi, S.A., Nedellec, R., Daniels, B.R., Mullins, J.I., Mosier, D.E., Siliciano, R.F., Wirtz, D., 2013. Adhesion and fusion efficiencies of human immunodeficiency virus type 1 (HIV-1) surface proteins. Sci. Rep. **3**, 3014.

Gorry, P.R., Ancuta, P., 2011. Coreceptors and HIV-1 pathogenesis. Curr. HIV/AIDS Rep. 8, 45–53.

Guenaga, J., Dubrovskaya, V., de Val, N., Sharma, S.K., Carrette, B., Ward, A.B., Wyatt, R.T., 2015. Structure-guided redesign increases the propensity of HIV env to generate highly stable soluble trimers. J. Virol. 90, 2806–2817.

Gulick, R.M., Su, Z., Flexner, C., Hughes, M.D., Skolnik, P.R., Wilkin, T.J., Gross, R., Krambrink, A., Coakley, E., Greaves, W.L., Zolopa, A., Reichman, R., Godfrey, C., Hirsch, M., Kuritzkes, D.R., Team ACTG, 2007. Phase 2 study of the safety and efficacy of vicriviroc, a CCR5 inhibitor, in HIV-1-Infected, treatment-experienced patients: AIDS clinical trials group 5211. J. Infect. Dis. 196, 304–312.

He, J., Choe, S., Walker, R., Di Marzio, P., Morgan, D.O., Landau, N.R., 1995. Human immunodeficiency virus type 1 viral protein R (Vpr) arrests cells in the G2 phase of the cell cycle by inhibiting p34cdc2 activity. J. Virol. 69, 6705–6711.

Henrich, T.J., Kuritzkes, D.R., 2013. HIV-1 entry inhibitors: recent development and clinical use. Curr. Opin. Virol. 3, 51–57.
Hong, P.W., Nguyen, S., Young, S., Su, S.V., Lee, B., 2007. Identification of the optimal DC-SIGN binding site on human immunodeficiency virus type 1 gp120. J. Virol. 81, 8325–8336.
Jensen, M.A., van't Wout, A.B., 2003. Predicting HIV-1 coreceptor usage with sequence analysis. AIDS Rev. 5, 104–112.
Jensen, M.A., Li, F.S., van't Wout, A.B., Nickle, D.C., Shriner, D., He, H.X., McLaughlin, S., Shankarappa, R., Margolick, J.B., Mullins, J.I., 2003. Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loop sequences. J . Virol. 77, 13376–13388.
Kirchherr, J.L., Lu, X., Kasongo, W., Chalwe, V., Mwananyanda, L., Musonda, R.M., Xia, S.M., Scearce, R.M., Liao, H.X., Montefiori, D.C., Haynes, B.F., Gao, F., 2007. High throughput functional analysis of HIV-1 env genes without cloning. J. Virol. Methods 143, 104–111.
Korber, B., Gnanakaran, S., 2009. The implications of patterns in HIV diversity for neutralizing antibody induction and susceptibility. Curr. Opin. HIV AIDS 4, 408–417.
Kwong, P.D., Wyatt, R., Robinson, J., Sweet, R.W., Sodroski, J., Hendrickson, W.A., 1998. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature 393, 648–659.
Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359.
Lengauer, T., Sander, O., Sierra, S., Thielen, A., Kaiser, R., 2007. Bioinformatics prediction of HIV coreceptor usage. Nat. Biotechnol. 25, 1407–1410.
Leonard, C.K., Spellman, M.W., Riddle, L., Harris, R.J., Thomas, J.N., Gregory, T.J., 1990. Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. J. Biol. Chem. 265, 10373–10382.
Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.
Low, A.J., McGovern, R.A., Harrigan, P.R., 2009. Trofile HIV co-receptor usage assay. Expert Opin. Med. Diagn. 3, 181–191.
Masso, M., Vaisman, I.I., 2010. Accurate and efficient gp120 V3 loop structure based models for the determination of HIV-1 co-receptor usage. BMC Bioinform. 11, 494.
Michael, N., Kim, J.H., 1999. HIV Protocols. Humana Press, Totowa, N.J.
Moore, J.P., Kuritzkes, D.R., 2009. A piece de resistance: how HIV-1 escapes small molecule CCR5 inhibitors. Curr. Opin. HIV AIDS 4, 118–124.
Moreno-Gamez, S., Hill, A.L., Rosenbloom, D.I., Petrov, D.A., Nowak, M.A., Pennings, P.S., 2015. Imperfect drug penetration leads to spatial monotherapy and rapid evolution of multidrug resistance. Proc. Natl. Acad. Sci. USA 112.
Pancera, M., Zhou, T., Druz, A., Georgiev, I.S., Soto, C., Gorman, J., Huang, J., Acharya, P., Chuang, G.Y., Ofek, G., Stewart-Jones, G.B., Stuckey, J., Bailer, R.T., Joyce, M.G., Louder, M.K., Tumba, N., Yang, Y., Zhang, B., Cohen, M.S., Haynes, B.F., Mascola, J. R., Morris, L., Munro, J.B., Blanchard, S.C., Mothes, W., Connors, M., Kwong, P.D., 2014. Structure and immune recognition of trimeric pre-fusion HIV-1 Env. Nature 514, 455–461.
Pastore, C., Nedellec, R., Ramos, A., Pontow, S., Ratner, L., Mosier, D.E., 2006. Human immunodeficiency virus type 1 coreceptor switching: V1/V2 gain-of-fitness mutations compensate for V3 loss-of-fitness mutations. J. Virol. 80, 750–758.
Picard, L., Simmons, G., Power, C.A., Meyer, A., Weiss, R.A., Clapham, P.R., 1997. Multiple extracellular domains of CCR-5 contribute to human immunodeficiency virus type 1 entry and fusion. J. Virol. 71, 5003–5011.
Poignard, P., Saphire, E.O., Parren, P.W., Burton, D.R., 2001. gp120: Biologic aspects of structural features. Annu. Rev. Immunol. 19, 253–274.
Pontow, S., Ratner, L., 2001. Evidence for common structural determinants of human immunodeficiency virus type 1 coreceptor activity provided through functional analysis of CCR5/CXCR4 chimeric coreceptors. J. Virol. 75, 11503–11514.
Poon, A.F., Swenson, L.C., Bunnik, E.M., Edo-Matas, D., Schuitemaker, H., van't Wout, A.B., Harrigan, P.R., 2012. Reconstructing the dynamics of HIV evolution within hosts from serial deep sequence data. PLoS Comput. Biol. 8, e1002753.
Prosperi, M.C., Bracciale, L., Fabbiani, M., Di Giambenedetto, S., Razzolini, F., Meini, G., Colafigli, M., Marzocchetti, A., Cauda, R., Zazzi, M., De Luca, A., 2010. Comparative determination of HIV-1 co-receptor tropism by enhanced sensitivity trofile, gp120 V3-loop RNA and DNA genotyping. Retrovirology 7, 56.
Ray, N., Doms, R.W., 2006. HIV-1 coreceptors and their inhibitors. Curr. Top. Microbiol. Immunol. 303, 97–120.
Rodriguez, C., Soulie, C., Marcelin, A.G., Calvez, V., Descamps, D., Charpentier, C., Flandre, P., Recordon-Pinson, P., Bellecave, P., Pawlotsky, J.M., Masquelier, B., AAS Group, 2015. HIV-1 coreceptor usage assessment by ultra-deep pyrosequencing and response to maraviroc. PLoS One 10, e0127816.
Rothenberger, M.K., Keele, B.F., Wietgrefe, S.W., Fletcher, C.V., Beilman, G.J., Chipman, J.G., Khoruts, A., Estes, J.D., Anderson, J., Callisto, S.P., Schmidt, T.E., Thorkelson, A., Reilly, C., Perkey, K., Reimann, T.G., Utay, N.S., Nganou Makamdop, K., Stevenson, M., Douek, D.C., Haase, A.T., Schacker, T.W., 2015. Large number of rebounding/founder HIV variants emerge from multifocal infection in lymphatic tissues after treatment interruption. Proc. Natl. Acad. Sci. USA 112, E1126–E1134.
Saita, Y., Kodama, E., Orita, M., Kondo, M., Miyazaki, T., Sudo, K., Kajiwara, K., Matsuoka, M., Shimizu, Y., 2006. Structural basis for the interaction of CCR5 with a small molecule, functionally selective CCR5 agonist. J. Immunol. 177, 3116–3122.
Salazar-Gonzalez, J.F., Bailes, E., Pham, K.T., Salazar, M.G., Guffey, M.B., Keele, B.F., Derdeyn, C.A., Farmer, P., Hunter, E., Allen, S., Manigart, O., Mulenga, J., Anderson, J.A., Swanstrom, R., Haynes, B.F., Athreya, G.S., Korber, B.T., Sharp, P. M., Shaw, G.M., Hahn, B.H., 2008. Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. J. Virol. 82, 3952–3970.
Samson, M., LaRosa, G., Libert, F., Paindavoine, P., Detheux, M., Vassart, G., Parmentier, M., 1997. The second extracellular loop of CCR5 is the major determinant of ligand specificity. J. Biol. Chem. 272, 24934–24941.
Schuitemaker, H., van't Wout, A.B., Lusso, P., 2011. Clinical significance of HIV-1 coreceptor usage. J. Transl. Med. 9 (Suppl. 1), S5.
Sede, M.M., Moretti, F.A., Laufer, N.L., Jones, L.R., Quarleri, J.F., 2014. HIV-1 tropism dynamics and phylogenetic analysis from longitudinal ultra-deep sequencing data of CCR5-and CXCR4-using variants. PLoS One 9, e102857.
Siciliano, S.J., Kuhmann, S.E., Weng, Y., Madani, N., Springer, M.S., Lineberger, J.E., Danzeisen, R., Miller, M.D., Kavanaugh, M.P., DeMartino, J.A., Kabat, D., 1999. A critical site in the core of the CCR5 chemokine receptor required for binding and infectivity of human immunodeficiency virus type 1. J. Biol. Chem. 274, 1905–1913.
Sirois, S., Sing, T., Chou, K.C., 2005. HIV-1 gp120 V3 loop for structure-based drug design. Curr. Protein Pept. Sci. 6, 413–422.
Smyth, R.P., Schlub, T.E., Grimm, A., Venturi, V., Chopra, A., Mallal, S., Davenport, M. P., Mak, J., 2010. Reducing chimera formation during PCR amplification to ensure accurate genotyping. Gene 469, 45–51.
Swenson, L.C., Mo, T., Dong, W.W., Zhong, X., Woods, C.K., Jensen, M.A., Thielen, A., Chapman, D., Lewis, M., James, I., Heera, J., Valdez, H., Harrigan, P.R., 2011. Deep sequencing to infer HIV-1 co-receptor usage: application to three clinical trials of maraviroc in treatment-experienced patients. J. Infect. Dis. 203, 237–245.
Tan, Q., Zhu, Y., Li, J., Chen, Z., Han, G.W., Kufareva, I., Li, T., Ma, L., Fenalti, G., Li, J., Zhang, W., Xie, X., Yang, H., Jiang, H., Cherezov, V., Liu, H., Stevens, R.C., Zhao, Q., Wu, B., 2013. Structure of the CCR5 chemokine receptor-HIV entry inhibitor maraviroc complex. Science 341, 1387–1390.
Thielen, A., Lengauer, T., Swenson, L.C., Dong, W.W., McGovern, R.A., Lewis, M., James, I., Heera, J., Valdez, H., Harrigan, P.R., 2011. Mutations in gp41 are correlated with coreceptor tropism but do not improve prediction methods substantially. Antivir. Ther. 16, 319–328.
Tsibris, A.M., Korber, B., Arnaout, R., Russ, C., Lo, C.C., Leitner, T., Gaschen, B., Theiler, J., Paredes, R., Su, Z., Hughes, M.D., Gulick, R.M., Greaves, W., Coakley, E., Flexner, C., Nusbaum, C., Kuritzkes, D.R., 2009. Quantitative deep sequencing reveals dynamic HIV-1 escape and large population shifts during CCR5 antagonist therapy in vivo. PLoS One 4, e5683.
Westervelt, P., Trowbridge, D.B., Epstein, L.G., Blumberg, B.M., Li, Y., Hahn, B.H., Shaw, G.M., Price, R.W., Ratner, L., 1992. Macrophage tropism determinants of human immunodeficiency virus type 1 in vivo. J. Virol. 66, 2577–2582.
Wilen, C.B., Tilton, J.C., Doms, R.W., 2012. HIV: cell binding and entry. Cold Spring Harb. Perspect. Med., 2.
Wu, B., Chien, E.Y., Mol, C.D., Fenalti, G., Liu, W., Katritch, V., Abagyan, R., Brooun, A., Wells, P., Bi, F.C., Hamel, D.J., Kuhn, P., Handel, T.M., Cherezov, V., Stevens, R.C., 2010. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. Science 330, 1066–1071.
Zhou, T., Georgiev, I., Wu, X., Yang, Z.Y., Dai, K., Finzi, A., Kwon, Y.D., Scheid, J.F., Shi, W., Xu, L., Yang, Y., Zhu, J., Nussenzweig, M.C., Sodroski, J., Shapiro, L., Nabel, G.J., Mascola, J.R., Kwong, P.D., 2010. Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. Science 329, 811–817.